

Journal of Educational Psychology

WAYNE H. HOLTZMAN, EDITOR
University of Texas



Advisory Editors

ROBERT H. BEEZER
United States Office of Education

DANIEL E. BERLYNE
University of Toronto

C. VICTOR BUNDERSON
University of Texas

LEE J. CRONBACH
Stanford University

NATHANIEL GAGE
Stanford University

ROBERT M. GAGNÉ
University of California, Berkeley

JAMES J. GALLAGHER
Duke University

ROBERT GLASER
University of Pittsburgh

ERNEST R. HILGARD
Stanford University

PHILIP W. JACKSON
University of Chicago

ARTHUR R. JENSEN
Center for Advanced Study in the Behavioral Sciences, Stanford, California

HERBERT J. KLAUSMEIER
University of Wisconsin

NATHAN KOGAN
Educational Testing Service, Princeton

ARTHUR A. LUMSDAINE
University of Washington

CARSON MCGUIRE
University of Texas

PAUL MUSSEN
University of California, Berkeley

C. ROBERT PACE
University of California, Los Angeles

DAVID G. RYANS
University of Hawaii

ROBERT L. THORNDIKE
Teachers College, Columbia University

BENTON J. UNDERWOOD
Northwestern University

VOLUME 58, 1967

HELEN ORR
Managing Editor

HENRY S. SALMON
Advertising and Business Manager

ELIZABETH MULLEN
Technical Editor

ROBERT D. GORMAN
Subscription Manager

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth Street, N. W.
WASHINGTON, D. C. 20036

Copyright, 1967, by the American Psychological Association, Inc.

Bureau of Ednl. & Psyl. Research
(S. C. E. R. I.) 6.3.81
Date
Acc. 7678

\$10.50

CONTENTS OF VOLUME 58

ALTER, RICHARD D. See WHITE, B. JACK	
ANDERSON, RICHARD C. See DUELL, ORPHA K.	
ANDERSON, RICHARD C. See FAUST, GERALD W.	
BAKER, ROBERT L. See SULLIVAN, HOWARD J.	
BAUMEISTER, ALFRED A., AND GUFFIN, JUDITH. Effects of Conceptual Similarity on Serial Learning and Retention by Retardates.....	308
BIRNEY, ROBERT C. See RIDLEY, DENNIS R.	
BISBICOS, ETHEL E., See ROTHKOPF, ERNST Z.	
BRAMBLE, WILLIAM J. See HORN, JOHN L.	
CLARK, VIRGINIA A. See KERRICK, JEAN S.	
COOPER, J. C., JR., AND GAETH, J. H. Interactions of Modality with Age and Meaningfulness in Verbal Learning.....	41
CROWELL, DORIS C. See FARGO, GEORGE A.	
DAW, ROBERT W., AND GAGE, N. L. Effect of Feedback from Teachers to Principals....	181
DENO, STANLEY L. See JENKINS, JOSEPH R.	
DI VESTA, FRANCIS J., AND WALLS, RICHARD T. Transfer of Solution Rules in Problem Solving.....	319
DOHERTY, MARY AUSTIN. See LAUGHLIN, PATRICK R.	
DUELL, ORPHA K., AND ANDERSON, RICHARD C. Pitch Discrimination among Primary School Children.....	315
DUERFELDT, PRYSE H. See KANFER, FREDERICK H.	
DUNN-RANKIN, PETER. See FARGO, GEORGE A.	
FAUST, GERALD W., AND ANDERSON, RICHARD C. Effects of Incidental Material in a Programmed Russian Vocabulary Lesson.....	3
FARGO, GEORGE A., CROWELL, DORIS C., NOYES, MARY H., FUCHIGAMI, ROBERT Y., GORDON, JOHN M., AND DUNN-RANKIN, PETER. Comparability of Group Television and Individual Administration of the Peabody Picture Vocabulary Test: Implications for Screening.....	137
FISHER, RHODA LEE. Social Schema of Normal and Disturbed School Children.....	88
FRASE, LAWRENCE T. Learning from Prose Material: Length of Passage, Knowledge of Results, and Position of Questions.....	266
FUCHIGAMI, ROBERT Y. See FARGO, GEORGE A.	
GAETH, J. H. See COOPER, J. C., JR.	
GAGE, N. L. See DAW, ROBERT W.	
GORDON, JOHN M. See FARGO, GEORGE A.	
GUFFIN, JUDITH. See BAUMEISTER, ALFRED A.	
GUTHRIE, JOHN T. Expository Instruction versus a Discovery Method.....	45
HAINES, DONALD BRUCE, AND MCKEACHIE, W. J. Cooperative versus Competitive Discussion Methods in Teaching Introductory Psychology.....	386
HAMERLYNCK, LEO A. See KNOWLTON, JAMES Q.	
HARRIS, CHESTER W. See LEMKE, ELMER A.	
HOLLAND, JOHN L. See RICHARDS, JAMES M., JR.	
HOLTZMAN, WAYNE H. Editorial.....	1
HORN, JOHN L., AND BRAMBLE, WILLIAM J. Second-Order Ability Structure Revealed in Rights and Wrongs Scores.....	115
HOROWITZ, HERBERT. Prediction of Adolescent Popularity and Rejection from Achievement and Interest Tests.....	170
HUMPHREYS, LLOYD G. Critique of Cattell's "Theory of Fluid and Crystallized Intelligence: A Critical Experiment".....	129
JACKSON, PHILIP W., AND LAHADERNE, HENRIETTE M. Scholastic Success and Attitude toward School in a Population of Sixth Graders.....	15
JENKINS, JOSEPH R., NEALE, DANIEL C., AND DENO, STANLEY L. Differential Memory for Picture and Word Stimuli.....	303

JOHNSON, PAUL E. Some Psychological Aspects of Subject-Matter Structure.....	75
KANFER, FREDERICK H., AND DUEFELDT, PRYSE H. Learner Competence, Model Competence, and Number of Observation Trials in Vicarious Learning.....	153
KARRAKER, R. J. Knowledge of Results and Incorrect Recall of Plausible Multiple-Choice Alternatives.....	11
KERRICK, JEAN S., CLARK, VIRGINIA A., AND RICE, DONALD T. Lecture versus Participation in the Health Training of Peace Corps Volunteers.....	259
KLAUSMEIER, HERBERT J. See LEMKE, ELMER A.	
KNOWLTON, JAMES Q., AND HAMERLYNCK, LEO A. Perception of Deviant Behavior: A Study of Cheating.....	379
LAHADERNE, HENRIETTE M. See JACKSON, PHILIP W.	
LAMB, GEORGE S. Teacher Verbal Cues and Pupil Performance on a Group Reading Test.....	332
LANE, PAUL A. See TOLOR, ALEXANDER.	
LAUGHLIN, PATRICK R., AND DOHERTY, MARY AUSTIN. Discussion versus Memory in Cooperative Group Concept Attainment.....	123
LEIBLUM, SANDRA L. See WHIMBEY, ARTHUR E.	
LEMKE, ELMER A., KLAUSMEIER, HERBERT J., AND HARRIS, CHESTER W. Relation of Selected Cognitive Abilities to Concept Attainment and Information Processing...	27
LEVIN, JOEL R. See ROHWER, WILLIAM D., JR.	
LOVITT, THOMAS C. Free-Operant Preference for One of Two Stories: A Methodological Note.....	84
LUTZ, SANDRA W. See RICHARDS, JAMES M., JR.	
LYNCH, STEVE. See ROHWER, WILLIAM D., JR.	
MCKEACHIE, W. J. See HAINES, DONALD BRUCE.	
MELARAGNO, RALPH J. Two Methods for Adapting Self-Instructional Materials to Individual Differences.....	327
MINIUM, EDWARD W. See PLANT, WALTER T.	
NEALE, DANIEL C. See JENKINS, JOSEPH R.	
NOYES, MARY H. See FARGO, GEORGE A.	
OVERING, ROBERT L. R., AND TRAVERS, ROBERT M. W. Variation in the Amount of Irrelevant Cues in Training and Test Conditions and the Effect upon Transfer.....	62
PARADOWSKI, WILLIAM. Effect of Curiosity on Incidental Learning.....	50
PERVIN, LAWRENCE A. A Twenty-College Study of Student X College Interaction Using TAPE (Transactional Analysis of Personality and Environment): Rationale, Reliability, and Validity.....	290
PLANT, WALTER T., AND MINIUM, EDWARD W. Differential Personality Development in Young Adults of Markedly Different Aptitude Levels.....	141
RICE, DONALD T. See KERRICK, JEAN S.	
RICHARDS, JAMES M., JR., HOLLAND, JOHN L., AND LUTZ, SANDRA W. Prediction of Student Accomplishment in College.....	343
RIDLEY, DENNIS, R., AND BIRNEY, ROBERT C. Effects of Training Procedures on Creativity Test Scores.....	158
RINGNESS, THOMAS A. Identification Patterns, Motivation, and School Achievement of Bright Junior High School Boys.....	93
ROHWER, WILLIAM D., JR., LYNCH, STEVE, LEVIN, JOEL R., AND SUZUKI, NANCY. Pictorial and Verbal Factors in the Efficient Learning of Paired Associates.....	278
ROTHKOPF, ERNST Z., AND BISBICOS, ETHEL E. Selective Facilitative Effects of Interspersed Questions on Learning from Written Materials.....	56
SAMUELS, S. JAY. Attentional Process in Reading: The Effect of Pictures on the Acquisition of Reading Responses.....	337
SASSENATH, JULIUS M., AND YONGE, GEORGE D. Meaning and Transfer of Verbal Learning.....	365
SCANDURA, JOSEPH M. Learning Verbal and Symbolic Statements of Mathematical Rules.....	356
SCARPETTI, WILLIAM L. See TOLOR, ALEXANDER.	
SCHUTZ, RICHARD E. See SULLIVAN, HOWARD J.	
SOTO, DOLORES H. See STANNERS, ROBERT F.	

STANNERS, ROBERT F., AND SOTO, DOLORES H. Developmental Changes in the Recognition of Beginning Segments of English Words.....	273
SULLIVAN, HOWARD J., BAKER, ROBERT L., AND SCHUTZ, RICHARD E. Effect of Intrinsic and Extrinsic Reinforcement Contingencies on Learner Performance.....	165
SUZUKI, NANCY. See ROHWER, WILLIAM D., JR.	
TAGATZ, GLENN E. Effects of Strategy, Sex, and Age on Conceptual Behavior of Elementary School Children.....	103
THALBERG, STANTON P. Reading Rate and Immediate versus Delayed Retention....	373
TOLOR, ALEXANDER, SCARPETTI, WILLIAM L., AND LANE, PAUL A. Teachers' Attitudes toward Children's Behavior Revisited.....	175
TRAVERS, ROBERT M. W. See OVERING, ROBERT L. R.	
WALLACE, WILLIAM P. Implicit Associative Response Occurrence in Learning with Retarded Subjects: A Supplementary Report.....	110
WALLS, RICHARD T. See DI VESTA, FRANCIS J.	
WHIMBEY, ARTHUR E., AND LEIBLUM, SANDRA L. Individual Differences in Memory Span with and without Activity Intervening between Presentation and Recall.....	311
WHITE, B. JACK, AND ALTER, RICHARD D. Dogmatism and Examination Performance...	285
WITTRICK, M. C. Replacement and Nonreplacement Strategies in Children's Problem Solving.....	69
WOLFF, JOSEPH LEE. Effects of Verbalization and Pretraining on Concept Attainment by Children in Two Mediation Categories.....	19
WOLFGANG, AARON. Effects of Social Cues and Task Complexity in Concept Identification.....	36
YONGE, GEORGE D. See SASSENATH, JULIUS M.	

WOOD, GORDON. Mnemonic Systems in Recall. *Journal of Educational Psychology Monograph*, Part 2, December 1967 (Whole No. 645). Pp. 1-27.

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 58

February 1967

Number 1

CONTENTS

<i>Editorial</i>	1
WAYNE H. HOLTZMAN	
<i>Effects of Incidental Material in a Programmed Russian Vocabulary Lesson</i>	3
GERALD W. FAUST AND RICHARD C. ANDERSON	
<i>Knowledge of Results and Incorrect Recall of Plausible Multiple-Choice Alternatives</i>	11
R. J. KARRAKER	
<i>Scholastic Success and Attitude toward School in a Population of Sixth Graders</i> ..	15
PHILIP W. JACKSON AND HENRIETTE M. LAHADERNE	
<i>Effects of Verbalization and Pretraining on Concept Attainment by Children in Two Mediation Categories</i>	19
JOSEPH LEE WOLFF	
<i>Relation of Selected Cognitive Abilities to Concept Attainment and Information Processing</i>	27
ELMER A. LEMKE, HERBERT J. KLAUSMEIER, AND CHESTER W. HARRIS	
<i>Effects of Social Cues and Task Complexity in Concept Identification</i>	36
AARON WOLFGANG	
<i>Interactions of Modality with Age and with Meaningfulness in Verbal Learning</i> ..	41
J. C. COOPER, JR. AND J. H. GAETH	
<i>Expository Instruction versus a Discovery Method</i>	45
JOHN T. GUTHRIE	
<i>Effect of Curiosity on Incidental Learning</i>	50
WILLIAM PARADOWSKI	
<i>Selective Facilitative Effects of Interspersed Questions on Learning from Written Materials</i>	56
ERNST Z. ROTHKOPF AND ETHEL E. BISBICOS	
<i>Variation in the Amount of Irrelevant Cues in Training and Test Conditions and the Effect upon Transfer</i>	62
ROBERT L. R. OVERING AND ROBERT M. W. TRAVERS	

Journal of Educational Psychology

EDITORIAL

During the past two years, education has "come alive" in ways that few would have predicted ten years ago. The tremendous upsurge of support for research and development, large-scale demonstrations under field conditions, the new reform movement in curriculum and educational technology, the advent of the computer and associated hardware will have a profound impact upon the field of educational psychology. In a similar manner, developmental psychology has been swept up in the vigorous new educational movement, resulting in greatly increased research and writing. Within the growing number of journals sponsored by the American Psychological Association, the *Journal of Educational Psychology* is emerging as the primary outlet for studies and theoretical articles in the closely related fields of educational and developmental psychology.

A relative newcomer to the group of journals published by APA, the *Journal* has grown appreciably in stature under the able leadership of Raymond Kuhlen, Editor during the first nine years of APA ownership. Kuhlen retires as Editor with the publication of this issue. Because of the usual publication lag, manuscripts submitted after May 1, 1966, will begin to appear in the April issue under the new Editor.

Although my term as Editor did not officially begin until January 1, 1967, responsibility for review of manuscripts and editorial policy was assumed in July in order to plan ahead for issues in 1967. During the four months from July to October, 124 manuscripts were received and are

now in various stages of processing. A sufficient number have been reviewed to permit some general impressions that may be helpful to authors. As in the past, only a small percentage of manuscripts are accepted, and many of these require some revision. A few manuscripts of good quality are rejected because they deal with topics outside the purview of the *Journal*. Much good educational research is of only minor psychological significance and is better suited for publication elsewhere. A major consideration in such decisions is the degree to which the problem is conceptualized in psychological terms and measures of psychological variables are employed.

Most rejected manuscripts are judged to be insufficient in quality or significance rather than outside the field. Serious defects in the formulation of the problem, the research design, the analysis or interpretation of results, or the general organization and presentation of a study are frequently uncovered by advisory editors to whom manuscripts are sent for critical review. Where the basic study is sound and the results potentially significant, detailed comments are sent to the author, who is encouraged to re-examine his data and revise his manuscript. In many cases, however, the shortcomings are too serious for revision to be of appreciable value. Other studies may be soundly executed and presented but are of minor importance. While some of these may be published as brief articles or notes, most simply fail to meet sufficient standards to justify space under present competition.

Preference will be given to contributions which have some relevance to education although such relevance is often hard to determine. Special efforts will be made to improve the

range and quality of contributions in developmental as well as educational psychology. Theoretical articles are especially welcome.

WAYNE H. HOLTZMAN, EDITOR

EFFECTS OF INCIDENTAL MATERIAL IN A PROGRAMMED RUSSIAN VOCABULARY LESSON

GERALD W. FAUST AND RICHARD C. ANDERSON

University of Illinois

In each of 2 experiments, 2 versions of a stylized program to teach Russian vocabulary were given to college students. 1 version used only primitive "copying frames," while the other version used copying frames in which the prompt sentence was embedded in a context of 5 English-Russian sentences. In both experiments the Context groups recalled significantly more ($p < .05$) overtly-practiced Russian words on the posttests than did the No-Context groups. The advantage of the Context programs was especially pronounced for Ss who hurried through the programs. The results can be interpreted as demonstrating that the addition of incidental material to the copying frame facilitates associative learning by insuring that Ss at least notice the stimulus before making the response.

The recent literature on programmed instruction has stressed the importance of getting the student to make the correct response for the right reasons. Holland (1965a) has expressed the point aptly, stating that "the answer required of the subject should be one he can give if, and only if, appropriate precursory behavior has occurred [p. 78]." "Appropriate precursory behavior" may include reading, examining a chart, or solving a problem. At a minimum, "appropriate precursory behavior" consists of noticing the stimulus that it is hoped will control the response in the future.

There is some evidence to support the principle that program effectiveness depends upon designing frames so that students cannot answer correctly unless they engage in the desired precursory behavior. Holland (1965b) and Krumboltz (1964) compared standard programs with programs modified so that trivial responses were required of subjects (Ss). In each case Ss who received the standard program did substantially better on the posttest than those who received the trivial-response program.

Maybe the trivial-response programs were poorer because Ss no longer had to engage in "appropriate precursory behavior" in order to respond correctly. But there is another explanation of these data. In the trivial-response programs Ss did not so often emit the important responses which they later would have to emit to earn a good score on the posttest. It is well known that making overt responses facilitates response-learning (McGeoch & Irion, 1952, p. 509), particularly when response m is low, such as is the case with unfamiliar, technical terms (Williams, 1965), foreign language vocabulary (Fry, 1960), or trigrams (Eigen & Margulies, 1963). It is quite possible that the trivial-word programs were inferior to the unaltered versions because in the altered programs Ss emitted important responses less often. Research that manipulates the likelihood of appropriate precursory behavior, while holding constant the amount of overt practice of important responses, is necessary to prove that program effectiveness depends upon insuring that the response is contingent upon appropriate precursory behavior.

Accompanying the argument that the response should be made contingent upon precursory behavior, such as reading relevant material, there is a second argument, treated as though it were a corollary: "Incidental material," that is, material upon which the response is *not* contingent, has no place in the self-instructional program. A high proportion of incidental material is claimed to be a sure sign of a bad program. The implication is that every shred of material irrelevant to the response in the frame in which the material appears should be excised.

The argument against incidental material does not follow logically from the principle that the response should be contingent upon appropriate precursory behavior. Provided a frame contains at least some relevant material upon which the response does depend, it would, as far as this principle is concerned, seem to be a matter of indifference as to whether the frame also contained incidental material.

While it cannot be assumed that much will be learned from the incidental material itself (Eigen & Margulies, 1963), there are several reasons for supposing that, under certain circumstances, the presence of incidental material might facilitate learning from the material which does entail overt responding. First, there is the obvious point that suitable incidental material could make a program more interesting, and perhaps as a result keep the student working at the program such that he comes under the control of the contingencies which it embodies. Second, incidental material might force stimulus discrimination and response differentiation not required by a "stripped down" version of a program. The "blackout procedure" (Holland & Kemp, 1965), or a similar technique, could probably be applied to a program so as to make correct responding easier on some frames, easier because

fine discriminations were no longer required. As a result of removing incidental material, students might prove unable to make certain discriminations on the posttest. Third, speaking more speculatively, incidental material might serve as "filler" so as to give a distributed practice effect or a spaced review effect. Fourth, under some conditions, incidental material might increase the likelihood that the student will notice or pay attention to the stimulus before making a response.

The chief purpose of the two experiments described in the next section was to demonstrate that incidental material can facilitate learning from material upon which overt responses are conditional by requiring Ss to notice the stimulus before making the response. Two versions of stylized programs to teach Russian vocabulary were prepared. One program employed primitive "copying frames" (Markle, 1964), consisting of a prompt sentence with an English subject and its Russian equivalent as the predicate nominative, and the same sentence immediately below with the Russian word replaced by a blank. The copying frame assures correct responses but it permits the student to complete the blank correctly without attending to that part of the stimulus object which it is hoped will control the response in the future. The S who is given a series of copying frames, like the one shown in Figure 1, could simply copy the Russian word ("stohl") in the blank without ever reading the complete sentences. Such behavior may lead to good response learning but it will certainly not lead to any hookup between the English stimulus ("table") and the Russian response ("stohl"). In other words, the obvious copying frame does not guarantee that the English word will become a discriminative stimulus for its Russian equivalent.

The second program was identical

COPYING FRAME

A table is a stohl.



A table is a _____.

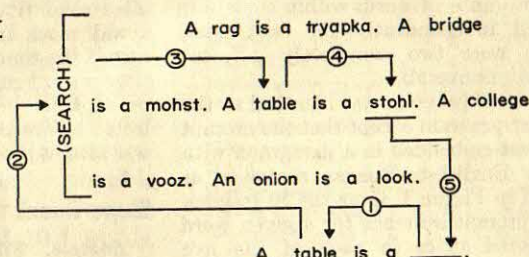
COPYING FRAME
WITH CONTEXT

FIG. 1. Two versions of the copying frame. (The lines trace the minimal eye movements that can be used to correctly fill in the response blank.)

to the first except that the prompt sentence was embedded in a paragraph with four other English-Russian sentences. Figure 1 illustrates the eye movements (a potentially observable aspect of "attention" or inspection behavior) necessary to complete a blank when incidental material is added to a copying frame. Notice that *S* must find the stimulus term and discriminate it from the surrounding material in order to fill in the blank. The student might not use the inspection technique diagrammed in Figure 1. He might read all of the sentences in the paragraph before making a response, or he might indulge in some covert rehearsal, and so on. However, the addition of incidental material guarantees that he will engage in at least minimally-adequate inspection behavior. Such frames are "failsafe" in the sense that *S* cannot simply copy a response without giving any attention to the stimulus. The prediction is that, despite possible interference from the incidental material, the addition of incidental material to the copying frame will result in a net increment in the learning of overtly-practiced items by facilitating S-R hookup.

METHOD

Experiment I

Sample. Forty-eight graduate students enrolled in an educational psychology course volunteered to act as *Ss* for this experiment. Students who had any knowledge of Russian were excluded from the sample.

Procedure. Each of the two groups received a training sequence entailing 10 presentations of 12 English-Russian word pairs, a posttest, new word pairs presented by the anticipation method and, finally, a retention test on the initially-learned words.

The Russian words used in the programs were selected from a list of 212 four- to seven-letter Russian words. Pronunciability ratings (PR) were obtained for each of these words from 34 summer institute participants, using the procedure described by Underwood and Schultz (1960). Two 12-word lists (List A and List B) were chosen from the 212 rated words such that each list contained six easy words (PR < 3.00) and six hard words (PR > 6.00). Each of these lists was further constrained to contain: (a) only words with common English meanings, (b) no two Russian words or English words with the same initial letter, (c) no two words with very similar phonetic or orthographic construction, (d) no Russian word with an obviously strong association to its English equivalent.

Each frame in the *No-Context* program consisted of two sentences. The first was the prompt sentence, containing an English subject and its Russian equivalent as the

predicate nominative. The second sentence was the same as the first except that the Russian word was replaced with a blank. This arrangement is pictured in Figure 1. Half of the Ss receiving the No-Context program were taught List A words. The remainder were taught List B words. The order of appearance of words within trials was randomized independently for each trial, and there were two completely different randomizations overall.

The *Context* program was identical to the No-Context program except that the prompt sentence was embedded in a paragraph with four other English-to-Russian sentences as is depicted in Figure 1. Over the 10 training trials the prompt sentence for a given word pair appeared twice in each of the five ordinal positions within the paragraph. Half of the Ss who received the Context program overtly practiced List A words and received List B words as part of the incidental material. The assignment was reversed for the rest of the Ss. Each incidental word on the unpracticed list appeared once per training trial, and over the 10 training trials the sentence containing any incidental word appeared twice at each ordinal position within the paragraph of five sentences. The remaining three English-Russian sentences can be called filler sentences to distinguish them from incidental sentences. The incidental sentences and filler sentences together comprised the context. Filler words were not systematically chosen. The order of frames was randomly determined for each trial independently of other trials. There were two completely different randomizations of the context materials.

The posttest and the retention test consisted of frames containing a sentence with an English subject and a blank in place of the Russian predicate nominative. Each test required recall of both overtly-practiced words and incidental words. Between the two tests all Ss received 10 anticipation trials on the incidental words. A strict scoring procedure, in which a Russian word was counted correct only if it was spelled correctly, was employed. A more relaxed scoring procedure was tried and then dropped since it did not appear that method of scoring interacted with other variables.

The eight sets of experimental materials (Context versus No-Context $\times$ Lists $\times$ Randomizations) were typed on buff-colored IBM cards and then placed in three-ring notebooks. Groups of between four and eight Ss completed the experiment at one time. Prior to each session, the materials

were stacked in a random order. After the Ss were seated, the materials were taken from this stack and passed out clockwise around the table. The Ss read the instructions which came with their programs and then were instructed by the experimenter on how to record the time both before and after each trial. All Ss were seated facing a wall clock from which they were able to record the time in minutes and seconds. The time record and all responses were written on separate answer sheets which were collected before the test trials. An experimenter was always present during sessions.

Experiment II

Sample. Fifty-four undergraduates enrolled in an educational psychology class acted as Ss to fulfill a course requirement. Once again Ss who were familiar with Russian were excluded from the sample.

Procedure. The two program versions used in this experiment were basically the same as those used in Experiment I. However, only one list of words received overt practice. The Ss were not tested for recall of incidental words. No retention test was given. The S received a posttest consisting of two test trials. In order to allow more chance for differentiation between program groups, lists of 16 words (8 hard words and 8 easy words) instead of the original 12 were used. In this experiment the PR of hard words ranged from 3.00 to 7.00. Because of the constraints imposed on word lists (see Experiment I) only 11 words from the original lists were used. The program format was changed slightly in that training and test frames were mimeographed and placed in consumable booklets. The Ss now wrote their responses into the program booklets instead of on separate answer sheets.

When Ss finished the posttest, they completed a questionnaire which inquired into attitudes toward the experiment and effort expended to learn the words. Six open-ended essay questions, presented first in the hope that Ss would volunteer information, asked for comment on the amount of rehearsal or self-testing, interest, persistence in attempting to learn the words, shortcuts used in filling in the blanks, etc. The Ss were then asked to complete nine multiple-choice questions intended to elicit the same information as the open-ended questions, but which were more pointed in wording.

In all other respects, the procedure in Experiment II was the same as that in Experiment I.

RESULTS

A one-tailed t test demonstrated the predicted superiority in recall of overtly-practiced words for the Context group in Experiment I ($t = 2.20$, $df = 46$, $p < .05$), though the analysis of variance (see Table 1) failed to show a significant main-effect difference for program. On the posttest and retention test combined, the Context group recalled a mean of 12.5 overtly-practiced words (highest possible score = 24) whereas the No-Context group recalled a mean of 10.4 words, a 20% advantage for the former group. This superiority can be traced primarily to the absence of very low scores in the Context group. The eight lowest scores were produced by Ss who received No-Context programs. Furthermore, the advantage of the Context group was due almost entirely to better recall of the easy words, as is evidenced by the significant interaction between Program $\times$ Word Difficulty (see Table 1). The Context group enjoyed a mean superiority of 2.1 easy words. There was no difference between the program groups in hard words recalled.

Still speaking of Experiment I, as had been expected, significantly fewer ($p < .01$) hard words than easy words were learned. Pooling over other conditions a mean of 2.8 hard words and a mean of 8.7 easy words were recalled. Significantly more ($p < .01$) words were recalled on the posttest than the retention test. Retention interval did not interact with any of the other variables investigated. The interaction of Word Difficulty $\times$ List was unexpectedly significant ($p < .01$), probably because the variability in word difficulty happened to be greater in List A.

A mean of .8 words from the incidental list was recalled on the posttest by the Context group, which, of course,

TABLE 1
ANALYSIS OF RECALL VARIANCE
IN EXPERIMENT 1

Source	df	MS	F
Between subjects			
Program	1	13.02	2.36
List (L)	1	.08	.002
Program $\times$ L	1	15.19	2.74
Persons within groups	44	5.55	
Within subjects			
Word difficulty (D)	1	426.00	161.36**
D $\times$ program	1	16.35	6.19*
D $\times$ L	1	38.17	14.16**
D $\times$ program $\times$ L	1	.69	.26
Persons within groups $\times$ D	44	2.64	
Retention Interval (R)	1	35.02	42.19**
R $\times$ program	1	.005	.0006
R $\times$ L	1	.17	.20
R $\times$ L $\times$ program	1	3.02	3.64
Persons within groups $\times$ R	44	.83	
D $\times$ R	1	1.35	2.33
D $\times$ R $\times$ program	1	.18	.31
D $\times$ R $\times$ L	1	1.13	1.94
D $\times$ R $\times$ program $\times$ L	1	1.62	2.79
Persons within groups $\times$ D $\times$ R	44	.58	

* $p < .05$.

** $p < .01$.

was the only group exposed to these words during training. Only five Ss recalled any incidental words. For these five a mean of 3.8 words was recalled and all but one of the incidental words recalled was an easy word. It should be emphasized that only one-fifth of the incidental material was tested. Extrapolating, probably an average of 4.0 incidental words was learned overall, accounted for entirely by a small percentage of Ss. Putting the matter another way, an S was about 15 times as likely to learn an overtly-practiced word as he was to learn an incidental word.

Dividing the Ss within each program group into quartiles on the basis of time taken to complete training, and then comparing the program groups on the basis of recall in each of these quartiles (Figure 2), reveals that the recall of the No-Context Ss was an increasing function of the amount of time spent on the training sequence.

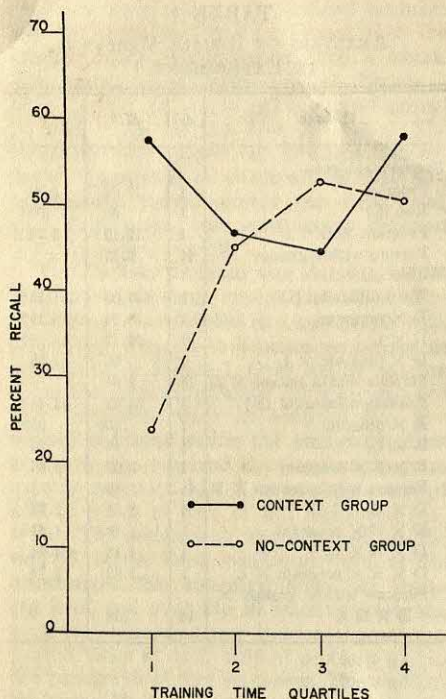


FIG. 2. Percentage of recall as a function of training time and program group in Experiment I.

This was not true of the Context Ss as there were only minor fluctuations in recall among their time quartiles. Analysis of variance showed that while neither the main effect of training time ($F = 2.53$, $df = 1/44$, $p > .05$) nor of program ($F = 2.75$, $df = 1/44$, $p > .05$) was significant, there was a significant interaction of Training Time $\times$ Program ($F = 8.02$, $df = 1/44$, $p < .01$). The interaction can be attributed to the superior performance of the Context Ss in the first time quartile. The Context group took longer to complete training ($t = 3.59$, $df = 46$, $p < .01$) than did the No-Context group.

Turning now to Experiment II, a one-tailed t test once again demonstrated the predicted superiority of the Context group in the recall of overtly-practiced words ($t = 1.98$, $df = 52$, p

$< .05$). On the two test trials the No-Context group recalled a mean of 20.0 words (highest possible score = 32) and the Context group recalled a mean of 21.7 words. Though the absolute difference in words recalled was only slightly less than in Experiment I, the Context group in Experiment II recalled only 9.1% more words than did the No-Context group, whereas in Experiment I the advantage was 20%. The percentage of difference was smaller in Experiment II because both groups learned a higher proportion of the words.

As in Experiment I the superiority of the Context group was due primarily to the absence of low recall scores. The Context group accounted for only 3 of the 14 cases in which 6 or fewer words were recalled.

Once again significantly more ($F = 89.5$, $df = 1/52$, $p < .01$) easy words than hard words were learned. A mean of 6.4 easy words and 4.1 hard words were recalled. The major difference between the groups was in the number of easy words recalled. The Context group enjoyed a mean advantage of 1.5 easy words and an advantage of only .2 hard words.

Figure 3, which compares the mean recall per training-time quartile for the two program groups, is very similar to Figure 2. In both Figure 3 and Figure 2 the Context group showed considerably greater recall in the first time quartile. In Figure 3, unlike Figure 2, this superiority was even greater in the second quartile. A one-tailed test of the interaction between Training Time $\times$ Program Condition showed the expected superiority of the Context group in the low time quartiles ($t = 2.25$, $df = 50$, $p < .05$).¹ The regression lines in Figure 3 show a steeper slope than those in Figure 2, a

¹ We are indebted to Gene V. Glass, personal communication, January, 1965, for developing the one-tailed interaction test.

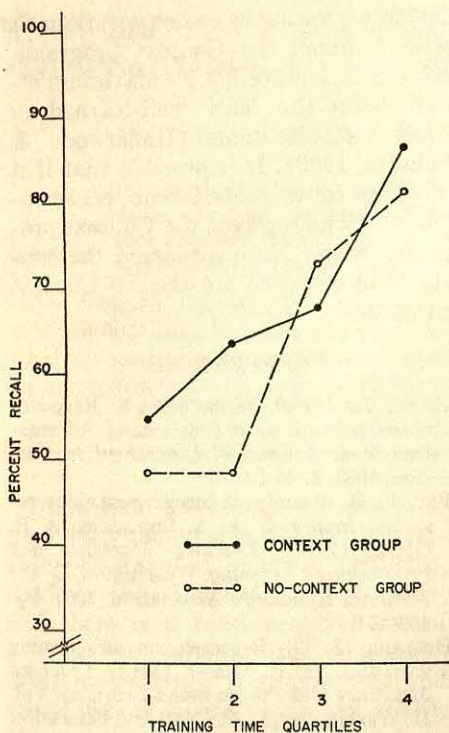


Fig. 3. Percentage of recall as a function of training time and program group in Experiment II.

fact that can probably be traced to the "easier hard words" used in Experiment II. The Ss who spent more time during training were able to make greater progress in learning the hard words. In other words, there was a higher ceiling than in Experiment I. The Context group took longer than the No-Context group to complete training ($t = 3.02$, $df = 52$, $p < .01$).

DISCUSSION

These experiments close one hole in the argument that the effectiveness of self-instructional programs depends upon designing frames so that "the answer required of the subject [is] one he can give if, and only if, appropriate precursory behavior has occurred [Holland, 1965a, p. 78]." Evidence previously advanced in support of this

principle, namely the studies involving trivial-response programs, could be explained in terms of a response-learning effect. In the present experiments all groups got the same amount of overt practice of important responses; hence, the facilitation can be attributed to the features of the Context programs that insured appropriate precursory behavior.

The results of the two experiments do support the contention that the addition of incidental material to the copying frame improves the performance of some Ss by requiring them to notice the discriminative stimulus before making the response. It should be emphasized that the prediction of superior recall by the Context groups assumes inadequate inspection behavior by the No-Context groups. Not all Ss in the No-Context groups employed inadequate inspection behavior. Only Ss who attempted to shortcut the blank-filling procedure by simply copying response terms were eliminating the chance for S-R hookup. Thus, only these Ss should be expected to fall short of their counterparts in the Context groups. Since Ss in the first and second time quartiles were those most likely to have been taking shortcuts in filling in the blanks, the superior performance of the first- and second-quartile Ss who used the Context programs, and the failure to find difference between programs for the third and fourth quartiles, is understandable.

The fact that Context groups showed an advantage on easy words but not on hard words is also understandable. Embedding the prompt sentence in a paragraph of incidental material literally requires S to "notice" the discriminative stimulus. Just noticing the discriminative stimulus may be a sufficient condition for associative learning in the case of easy words. Our belief is that just noticing the stimulus

was not sufficient in the case of hard words. For hard words *S* may need to search for mediating associations or engage in self-testing and covert rehearsal. There is little reason for believing that the Context treatment could "buy" these additional precursory behaviors.

The questionnaire data from the second experiment support the analysis of inspection behavior that has been advanced, particularly the interpretation of training time. Rated "conscientiousness" correlated .44 with training time. When compared to *Ss* with slower-than-median training times, those with faster-than-median times (both program groups pooled) reported a significantly greater tendency ($t = 2.07$, $df = 52$, $p < .05$) to "just try to fill in the blanks as quickly as possible [without spending] extra time studying the words." Many of those with fast training times actually reported using the inspection behaviors diagrammed in Figure 1.

The first experiment demonstrated once again that it is not safe to assume that *Ss* will learn from incidental material contained in programs.

There was some indirect evidence that the incidental material interfered to some extent with the recall of overtly-practiced words. Obviously interference could occur only for *Ss* who learned or partially learned incidental words. Four of the five *Ss* who recalled any incidental words were in the fourth training-time quartile. Considering now only *Ss* in the fourth quartile, those who received Context programs actually recalled fewer overtly-practiced hard words than comparable *Ss* who received No-Context programs. This fact is consistent with the notion that interference af-

fected performance on conscientious *Ss* who received the Context programs, since interference has its maximum effect upon the least well-learned or most difficult items (Underwood & Schultz, 1960). It is possible that if it were not for this interference conscientious *Ss* who received the Context programs would have surpassed the conscientious *Ss* who got the No-Context programs.

REFERENCES

- EIGEN, L. D., & MARGULIES, S. Response characteristics as a function of information level. *Journal of Programed Instruction*, 1963, 2, 45-54.
- FRY, E. B. A study of teaching-machine response modes. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching Machines and Programmed Learning*. Washington, D. C.: National Education Association, 1960. Pp. 469-474.
- HOLLAND, J. G. Research on programing variables. In R. Glaser (Ed.), *Teaching Machines and Programmed Learning*. Vol. II. Washington, D. C.: National Education Association, 1965. Pp. 66-117. (a)
- HOLLAND, J. G. Response contingencies in teaching-machine programs. *Journal of Programed Instruction*, 1965, 3, 1-8. (b)
- HOLLAND, J. G., & KEMP, F. D. A measure of programing in teaching-machine material. *Journal of Educational Psychology*, 1965, 56, 264-269.
- KRUMBOLTZ, J. D. The nature and importance of the required response in programed instruction. *American Educational Research Journal*, 1964, 4, 203-209.
- MARKLE, S. M. *Good frames and bad*. New York: Wiley, 1964.
- MCGEOCH, J., & IRION, A. *The psychology of human learning*. New York: Green, 1952.
- UNDERWOOD, B. J., & SCHULTZ, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.
- WILLIAMS, J. P. Effectiveness of constructed-response and multiple-choice programing modes as a function of test mode. *Journal of Educational Psychology*, 1965, 56, 111-117.

(Received March 22, 1966)

KNOWLEDGE OF RESULTS AND INCORRECT RECALL OF PLAUSIBLE MULTIPLE-CHOICE ALTERNATIVES¹

R. J. KARRAKER

University of Missouri at Kansas City

To test the hypothesis that plausible wrong responses in multiple-choice programs are recalled as being correct, 72 college freshmen were stratified into high- and low-ability levels, then randomly assigned to No Knowledge of Results, Knowledge of Results, or Control treatments. The Ss responded on a multiple-choice test in educational psychology, then a recall criterion test was administered. The score was the number of responses that were the same as the plausible wrong responses in the multiple-choice test. No Knowledge of Results resulted in significantly more ($p < .01$) errors than Knowledge of Results or Control. When No Knowledge of Results was given, Ss did recall more plausible wrong responses as being correct, but this effect did not appear when Knowledge of Results was given.

Immediate knowledge of results is frequently included in lists of dependable principles of learning (Hilgard, 1956, p. 487; Watson, 1960, p. 254), and there is a considerable body of evidence to validate this conclusion using many types of apparatuses and species of organisms (Ammons, 1956). However, in applying this principle to verbal learning, the issue has been found to be less axiomatic. For instance, two investigators report no significant difference attributable to knowledge of results (Feldhusen & Birt, 1962; Hough & Revsin, 1963), and others report knowledge of results to be interacting with level of achievement and difficulty of item (Goldbeck & Campbell, 1962; Gordon, 1966); motivation (Kight & Sassenrath, 1966); and duration of the response-knowledge of results interval (Evans, 1960).

However, Pressey (1950) and Angell and Troyer (1948) report unqualified

advantages in utilizing a machine and a device called the punchboard to provide immediate knowledge of results to improve verbal learning. Skinner also considers knowledge of results an extremely important variable in the learning process. Skinner (1954, p. 91) states, "... the lapse of only a few seconds between response and reinforcement (knowledge of successful results) destroys most of the effect."

Skinner considers it desirable for subjects (Ss) to receive knowledge of results in order for reinforcement to occur. Consequently, Skinner (1958, p. 970) suggests that verbal material must always be written to "... help the student to come up with the right answer." Pressey (1960, p. 501) considers it desirable that students be exposed to incorrect alternatives to teach frequently occurring misunderstandings. Skinner (1958, p. 970) states, "... effective multiple-choice material must contain plausible wrong responses, which are out of place in the delicate process of shaping behavior because they strengthen unwanted forms."

Kaess and Zeaman (1960) report that negative knowledge of results (knowledge of unsuccessful results)

¹Part of the data was presented at the February, 1966, meeting of the American Educational Research Association, Chicago, Illinois. Appreciation is expressed to K. D. Orton and Clinton I. Chase for consultative services during the preparation of this paper.

has a deleterious effect on learning. They administered a multiple-choice test, gave knowledge of results, and readministered the same test. They found that students responded more often than chance would suggest to their prime distractor, indicating some students learned the wrong response from exposure to a wrong alternative. However, of the responses that were incorrect on the initial trial, 53% were correct on the second trial.

Pressey (1950) found that after administration of a multiple-choice test with a punchboard, students showed a gain in number of correct responses on a retake of the same test, on transfer tests, and on tests requiring a constructed response. However, there was no information given in regard to the extent to which students responded on the transfer and constructed response tests with incorrect alternatives recalled from the multiple-choice test. The latter seems more directly to answer the question; Do plausible-sounding wrong alternatives strengthen unwanted forms of behavior?

METHOD

Subjects

The Ss were 72 college freshmen enrolled in a class entitled Introduction to Educational Psychology. They were divided into ability levels on the basis of their "Gamma IQ" on the Otis Quick-Scoring Mental Ability Test. The mean of the high-ability group was 130 and that of the low-ability group was 121. The Ss were randomly assigned to treatments within their ability levels.

Treatments

Knowledge of Results. The Ss were administered a multiple-choice achievement test in educational psychology. The following class period, the test was again distributed among Ss and the experimenter (*E*) read the items with the correct answers. No discussion of the items was permitted. The class was told there was not sufficient time available for discussion. The *E* did not give any indication that a constructed response

test would be administered the following class period.

No Knowledge of Results. The Ss were administered the same multiple-choice test, but were not given access to the test again.

Control. The Ss did not take the multiple-choice test.

Tests Employed

The 40 items in the multiple-choice test covered a number of topics in educational psychology. Reliability of the test estimated by the Spearman-Brown prophecy formula was .84.

On the third class period, Ss in all treatment groups were administered the criterion test. This test was a constructed response test utilizing the same stem of the multiple-choice item, but requiring Ss to compose or recall the correct answer. Their score was the number of recall responses that were in essence the same as one of the three wrong alternatives in the multiple-choice test. The *E* scored these constructed responses "blind," in that he did not know to which treatment group individual Ss had been assigned.

When the criterion test was scored for correct answers, the reliability coefficient estimated by the Spearman-Brown prophecy formula was .78. The mean number of correct responses for the Knowledge of Results treatment was higher ($M = 34$, $SD = 3$) than the Control ($M = 28$, $SD = 7$) or the No Knowledge of Results ($M = 31$, $SD = 5$). In comparing these three means, only the Knowledge of Results and Control difference was significant at the .01 level ($t = 3.78$; $df = 46$; two-tailed test). The Knowledge of Results and No Knowledge of Results comparison yielded a t of 2.47, which was significant at the .05 level ($df = 46$; two-tailed test).

RESULTS

Means and standard deviations of the plausible wrong responses from the multiple-choice test that were recalled as being correct on the constructed response test are given in Table 1. The data were collapsed across levels for the treatment means, and across treatments for the ability means.

To test for differences among the groups, a 3 (treatments) $\times$ 2 (levels)

TABLE 1

MEAN AND STANDARD DEVIATION OF
PLAUSIBLE WRONG RESPONSES FROM
MULTIPLE-CHOICE TEST RECALLED
ON CONSTRUCTED RESPONSE TEST

Group	N	M	SD
Knowledge of results	24	3.54	2.53
No knowledge of results	24	6.33	2.82
Control	24	3.70	2.51
Low ability	36	5.19	3.14
High ability	36	3.86	2.50

factorial analysis of variance was used. An F of 8.93 ($df = 2/66$; $p < .001$) was obtained for treatments, and an F of 4.99 ($df = 1/66$; $p < .01$) was obtained for ability levels. The Treatment $\times$ Ability Levels interaction was not significant, as it gave an F of 1.49 ($df = 2/66$; $p > .05$).

The Duncan multiple-range test revealed the No Knowledge of Results treatment resulted in significantly more ($p < .01$) errors than the Knowledge of Results or Control treatments. The Knowledge of Results and Control treatments did not differ significantly. The low-ability group made significantly more ($p < .01$) errors than the high-ability group. The interaction was not significant.

DISCUSSION

The results of this study can be interpreted in at least two ways. If one attends to the fact that the most errors were made by the group that was exposed to the plausible wrong responses without knowledge of results, one concludes with Kaess and Zeaman (1960) that plausible wrong responses interfere with learning. In a recent review of the literature, Holland (1965, p. 104) arrives at the same conclusion.

However, all programs provide for knowledge of results, so the more important comparison becomes the Knowledge of Results and the Control

treatments. It seems noteworthy that in this experiment, when knowledge of results was given, the plausible wrong responses had no more effect than when Ss were never exposed to them. Consequently, from this study it appears that the rationale and experimental evidence on which linear programming is based need not necessarily exclude frames which utilize multiple-choice questions.

The nature of the terminal behavior desired may dictate the type of response mode. In the shaping of particular responses, the multiple-choice frame may be "out of place." As Schramm (1964, p. 10) has suggested, there are probably some learning tasks for which one or the other method works better, and Williams' (1963) distinction between items requiring technical terminology (where constructed response was superior) and items requiring general, familiar vocabulary is a case in point.

It should be mentioned that this multiple-choice test, and the verbal materials used by Kaess and Zeaman (1960) and Pressey (1950) are not programmed materials in that items are not arranged in logical sequence designed to lead the student in steps to the terminal behavior desired. Such materials have been termed "adjunct programming" (Holland, 1965, p. 67). The rationale for including these kinds of experiments in the field of programmed instruction has been developed elsewhere (Schramm, 1964, p. 1), but it is an obvious possibility that in other types of programming, results might differ.

REFERENCES

- AMMONS, R. B. Effect of knowledge of performance? A survey and tentative theoretical formulation. *Journal of Genetic Psychology*, 1956, **54**, 279-299.
- ANGELL, G. W., & TROYER, M. E. A new self-scoring device for improving instruction. *School and Society*, 1948, **56**, 84-85.

- EVANS, J. L. An investigation of teaching machine variables using learning programs in symbolic logic. Unpublished doctoral dissertation, University of Pittsburgh, 1960.
- FELDHUSEN, J. F., & BIRT, A. A study of nine methods of presentation of programmed learning material. *Journal of Educational Research*, 1962, **55**, 461-466.
- GOLDBECK, R. A., & CAMPBELL, V. N. The effects of response mode and response difficulty on programmed learning. *Journal of Educational Psychology*, 1962, **53**, 110-118.
- GORDON, J. M. Interaction effects of varying step size and feedback in programmed instruction. Paper read at American Educational Research Association, Chicago, February 1966.
- HILGARD, E. R. *Theories of learning*. New York: Appleton-Century-Crofts, 1956.
- HOLLAND, J. G. Research on programing variables. In R. Glaser (Ed.), *Teaching machines and programmed learning*. Vol. 2. *Data and directions*. Washington: National Education Association, 1965. Pp. 66-117.
- HOUGH, J. B., & REVSIN, B. Programed instruction at the college level: A study of several factors influencing learning. *Phi Delta Kappan*, 1963, **44**, 286-291.
- KAESS, W., & ZEAMAN, D. Positive and negative knowledge of results on a Pressey-type punchboard. *Journal of Experimental Psychology*, 1960, **60**, 12-17.
- KIGHT, H. R., & SASSENATH, J. M. Relation of achievement motivation and test anxiety to performance in programed instruction. *Journal of Educational Psychology*, 1966, **57**, 14-17.
- PRESSEY, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *Journal of Psychology*, 1950, **29**, 417-447.
- PRESSEY, S. L. Some perspectives and major problems regarding teaching machines. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning. A source book*. Washington: National Education Association, 1960. Pp. 497-506.
- SCHRAMM, W. *The research on programed instruction: An annotated bibliography*. Washington: United States Government Printing Office, 1964.
- SKINNER, B. F. The science of learning and the art of teaching. *Harvard Educational Review*, 1954, **24**, 86-97.
- SKINNER, B. F. Teaching machines. *Science*, 1958, **128**, 969-977.
- WATSON, G. What psychology can we feel sure about? *Teachers College Record*, 1960, **61**, 253-257.
- WILLIAMS, J. P. A comparison of several response modes in a review program. *Journal of Educational Psychology*, 1963, **54**, 253-260.

(Received January 26, 1966)

SCHOLASTIC SUCCESS AND ATTITUDE TOWARD SCHOOL IN A POPULATION OF SIXTH GRADERS¹

PHILIP W. JACKSON AND HENRIETTE M. LAHADERNE

University of Chicago

The study examined the accuracy of teachers' judgments of their students' satisfaction with school, and the relationship between scholastic success and attitude toward school. 292 6th graders responded to 2 questionnaires that assessed attitudes toward their schools and their teachers. IQ, achievement test scores, and course grades were obtained. Each teacher estimated each of his students' overall satisfaction with school. The results show that the teachers' estimations were better than chance, but they were related more closely to the students' academic record than to their expressed attitudes. The correlations between students' satisfaction scores and scholastic scores were negligible. There were no sex differences.

Success and satisfaction are bound together by logic, if not by fact. Logically at least, successful people ought to appear satisfied, and unsuccessful people dissatisfied, when queried about the conditions surrounding their achievements. In educational terms, students who are doing well in school might be expected to express contentment when asked to describe their school experience, and those who are doing poorly might be expected to express discontentment. Surprisingly, however, educational research has not yet provided a confirmation of this logically compelling expectation. Indeed, over the past 25 years an impressive amount of evidence has accumulated showing that scholastic success and attitudes toward school are typically unrelated to each other (Diedrich, 1966; Jackson & Getzels, 1959; Malpass, 1953; Tenenbaum, 1944; Tschechtelin, Hipskind, & Remmers, 1940).²

If confirmed by further investiga-

tion, the absence of a strong linkage between success and satisfaction in the classroom should provoke considerable thought among practitioners and researchers alike. Few of the questions to which it gives rise are more intriguing than those that concern the beliefs and behaviors of classroom teachers. How sensitive, for example, are teachers to differences in their students' views of school? When estimating the attitudes of their students do teachers act in accord with the popular expectation linking success and satisfaction or are their judgments in agreement with the results of empirical studies? The study reported here was designed to provide partial answers to these questions and to explore further the general relationship between scholastic success and attitudes toward school.

SUBJECTS AND PROCEDURES

The subjects (Ss) comprised the entire sixth grade of the public schools in a predominantly white, working class suburb (11 classes located in 6 schools; $N = 148$ boys, 144 girls). The pupils' mean IQ, as measured by the Kuhlmann-Anderson Intelligence Test, was 103.9, with a standard deviation of 19.2. The 11 sixth-grade teachers (7 women, 4 men) also participated in the study by estimating how satisfied each of the students was with his school experience.

¹ Expanded version of a paper presented at the American Educational Research Association meeting, Chicago, February 1966.

² One investigator (Brodie, 1964) does present evidence in support of the prediction, but his results, which seem to hold only for the girls in his sample, are difficult to evaluate because of the atypical performance of one of his experimental groups.

Toward the end of the school term, one of the investigators administered two attitude inventories to Ss during regular class periods. To encourage honesty in replying, Ss were assured that their responses would not be seen by their teachers nor by anyone else connected with the schools. The two inventories were:

1. The Student Opinion Poll II. This was a revision of a questionnaire developed a few years ago (Jackson & Getzels, 1959) and designed to elicit responses concerning general satisfaction or dissatisfaction with four aspects of school life: the teachers, the curriculum, the student body, and classroom procedures. The version used in this study contained 47 multiple-choice items and was scored by giving one point each time S chose from within a set of multiple-choices the response indicating the highest degree of satisfaction with that aspect of school life. Thus, the possible range of scores was from 0 to 47. The mean scores for the sample of sixth graders were 25.3 for boys and 29.4 for girls. The standard deviation for both sexes was 8.2. Test reliability for the total sample, based on Kuder-Richardson formula 20, was .86. The following are sample items.

6. The things that I am asked to study are of:

- a. great interest to me.
- b. average interest to me.
- c. little interest to me.
- d. no interest to me.

25. Teachers in this school seem to be:

- a. fair at all times.
- b. generally fair.
- c. occasionally unfair.
- d. often unfair.

47. In general, my feelings toward school are:

- a. very favorable—I like it as it is.
- b. somewhat favorable—I would like a few changes.
- c. somewhat unfavorable—I would like many changes.
- d. very unfavorable—I frequently feel that school is pretty much a waste of time.

2. The Michigan Student Questionnaire (abbreviated version). This was a shortened form of a questionnaire developed by Flanders and his associates (Flanders, 1965) to assess students' attitudes toward their present teachers and schoolwork. The version used in this study contained 37 descriptive statements, each followed by four possible replies: strongly disagree, disagree, agree, and strongly agree. A student's response to each item was scored 4, 3, 2, or 1 depending on the degree to which

his reply reflected a positive attitude toward his school and his teacher. Thus, the possible range of scores was from 37 to 148. The mean scores for the sample of sixth graders were 101.5 for boys and 109.3 for girls. Standard deviations were 19.2 and 16.9 for boys and girls respectively. Test reliability for the total sample, based on a variation of the Kuder-Richardson formula appropriate for weighted scores (Ferguson, 1951) was .94. The following are sample items.

12. What we learn in this class makes me want to learn new things.

Strongly disagree Disagree

Agree Strongly agree

16. This teacher certainly knows how to teach.

Strongly disagree Disagree

Agree Strongly agree

23. I really like this class.

Strongly disagree Disagree

Agree Strongly agree

The correlation between responses to the Student Opinion Poll II and the Michigan Student Questionnaire was .62 for the total sample. Thus, although the two questionnaires provide similar information, the relationship between them is low enough to justify the use of both in a study of students' attitudes.

At a special meeting held after school, the 11 sixth-grade teachers were shown sample items from the Student Opinion Poll II and were asked to predict how their students might respond to such a questionnaire. The exact procedure for obtaining the teachers' ratings was as follows. Each teacher was presented with an alphabetized list of his students. He was asked, first, to divide the group into thirds by classifying his students into three levels of satisfaction: "most," "average," and "least." He was then asked to identify from within the groups labeled "most" and "least" a smaller number of students (one-fourth of each group) who seemed to represent extreme positions ("very satisfied" and "very dissatisfied"). Thus, each student's attitudes were described by his teacher as falling into one of five categories. In each classroom the approximate fractions of students in the five categories were: $\frac{1}{12}$, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{12}$. When the ratings were treated quantitatively the values 15, 12, 10, 8, and 5 were assigned to the five groupings, the highest number being used to represent the attitudes of students whom the teachers had described as "very satisfied."

Two measures of scholastic performance were used. The first consisted of four grades, given by the S's present teacher, in reading,

TABLE 1
CORRELATIONS BETWEEN TEACHERS' ESTIMATES AND MEASURES OF STUDENT ATTITUDES,
ACHIEVEMENT, AND IQ

Students	Attitude measures		Achievement Measures			IQ
	Student Opinion Poll II	Michigan Student Questionnaire	Reading	Language	Arithmetic	
Boys	.28**	.27**	.49**	.51**	.45**	.44**
Girls	.27**	.25**	.36**	.37**	.31**	.39**

Note.— $N = 148$ for boys; $N = 144$ for girls.

** $p < .01$.

language arts, arithmetic, and science. The second consisted of three scores, derived from the Stanford Achievement Test, in reading, language arts, and arithmetic.

RESULTS

The findings with respect to the relationship between the teachers' estimates of students' satisfaction and scores derived from the students' responses to the two questionnaires are presented in Table 1. That table also contains correlations between the teachers' estimates and measures of the students' academic performance. The findings reveal that students' satisfaction is at least partially visible to teachers and can be estimated with greater-than-chance accuracy. They also reveal, however, that teachers tend to expect achievement and satisfaction to be more closely related than they, in fact, are. Indeed, the correlations in Table 1 indicate that when teachers set out to estimate how a student will respond to an attitude questionnaire, they come closer to describing how well the student achieves in school than to how he feels about his school experience.

Table 2 contains correlations between attitudes toward school and measures of scholastic achievement. Four features of that table deserve attention. First, and most important, all of the correlation coefficients are small. Second, they are of the same

magnitude with teachers' grades as with achievement test scores. Third, although the coefficients are uniformly higher for the Student Opinion Poll II than for the Michigan Student Questionnaire, the difference between the two sets of statistics is trivial. Fourth, there are no significant sex differences. Thus, these data confirm the results of other investigators who have found no significant relationship between attitudes toward school and scholastic achievement. Furthermore, they extend our previous knowledge by

TABLE 2
CORRELATIONS BETWEEN ATTITUDES AND
SCHOLASTIC PERFORMANCE

Scholastic performance	Attitudes			
	Student Opinion Poll II		Michigan Student Questionnaire	
	Boys	Girls	Boys	Girls
Grades				
Reading	.15	.16	.01	.06
Language	.13	.16	.01	.01
Arithmetic	.08	.14	.00	.00
Science	.15	.19*	.06	.04
Achievement tests				
Reading	.14	.08	.08	-.07
Language	.11	.14	.02	-.06
Arithmetic	.13	.12	.06	-.05
IQ	.06	.14	-.08	.01

Note.— $N = 148$ for boys; $N = 144$ for girls.

* $p < .05$.

demonstrating the absence of the success-satisfaction relationship with a wider variety of measuring instruments than has been used in earlier studies.

DISCUSSION

Several conditions might account for the unexpected lack of relationship between success and satisfaction. One possibility is that the range and intensity of student attitudes are not as great as responses to the questionnaire indicate and, thus, are not sufficiently powerful to affect behavior. Perhaps students typically do not either hate school or love it but, instead, feel rather neutral about their classroom experience. Another possibility is that teachers and parents behave in ways that effectively weaken whatever natural connection might exist between attitudes and achievement. In most classrooms students are required to master the minimal curricular objectives whether they want to or not. Assignments are clearly defined, deadlines are set, and frequent checks are made by teachers and parents alike to determine whether the work is being completed as expected.

Thus, in several ways teachers, parents, and general classroom conditions may counteract the natural consequences of differences in students' attitudes. The insignificant correlations between responses to questionnaires

and achievement test scores could be interpreted as quantitative evidence of the effectiveness of that counteraction.

REFERENCES

- BRODIE, T. A. Attitude toward school and academic achievement. *Personnel and Guidance Journal*, 1964, 43, 375-378.
- DIEDRICH, R. Teacher perceptions as related to teacher-student similarity and student satisfaction with school. Unpublished doctoral dissertation, University of Chicago, 1966.
- FERGUSON, G. A. A note on the Kuder-Richardson formula. *Educational and Psychological Measurements*, 1951, 11, 612-615.
- FLANDERS, N. A. Teacher influence, pupil attitudes, and achievement. *Cooperative Research Monograph*, No. 12. (OE-25040) Washington, D.C.: United States Government Printing Office, 1965.
- JACKSON, P. W., & GETZELS, J. W. Psychological health and classroom functioning: A study of dissatisfaction with school among adolescents. *Journal of Educational Psychology*, 1959, 50, 295-300.
- MALPASS, L. F. Some relationships between students' perceptions of school and their achievement. *Journal of Educational Psychology*, 1953, 44, 475-482.
- TENENBAUM, S. Uncontrolled expressions of children's attitudes toward school. *Elementary School Journal*, 1940, 40, 670-678.
- TENENBAUM, S. Attitudes of elementary school children to school, teachers, and classmates. *Journal of Applied Psychology*, 1944, 28, 134-141.
- TSCHECHELIN, Sr. M. A., HIPSKIND, Sr. M. J. F., & REMMERS, H. H. Measuring the attitudes of elementary school children toward their teachers. *Journal of Educational Psychology*, 1940, 31, 195-203.

(Received February 18, 1966)

EFFECTS OF VERBALIZATION AND PRETRAINING ON CONCEPT ATTAINMENT BY CHILDREN IN TWO MEDIATION CATEGORIES¹

JOSEPH LEE WOLFF²

Institute of Educational Research, Indiana University

The study was concerned with the effects of 3 variables on concept attainment in an Osler-type concept-attainment task: mediational ability, verbalization, and pretraining. Both pretraining and verbalization had a significant positive effect on concept attainment. However, mediational ability, which was measured by means of a task developed by Kendler, Kendler, and Learnard, was not related to performance on the concept-attainment task. Since successful performance on this task requires that S use verbal mediators, it was concluded that the results of the study provide no support for the hypothesis that mediational ability as defined by the Kendler, Kendler, and Learnard task is a function of the presence or absence of verbal mediating processes.

Several two-stage S-R models have been used to predict human behavior in the concept-reversal-learning task. In the Kendler and Kendler (1962) model (hereafter the verbal mediation hypothesis) the hypothetical mediating response is assumed to be verbal or in some way verbally directed. House and Zeaman (1962; Zeaman & House, 1963), on the other hand, have postulated a purely perceptual mediator (dimensional orienting response), which serves simply to increase the probability that relevant cues will be observed.

According to the Kendlers, the two-stage model fits only the concept-reversal behavior of older children and adults (mediators); the behavior of younger children and rats (nonmediators), since they do not normally

mediate in solving such tasks, may be best described in accord with a traditional single-stage S-R model such as Spence's (1936). Greatly facilitating evaluation of the Kendlers' theory is a test technique (hereafter known as the mediation test), designed by Kendler, Kendler, and Learnard (1962) to identify children as either mediators or nonmediators. Thus, should children differentially categorized by the mediation test be given a motor discrimination task known a priori to be facilitated by *verbal* mediation, it could be predicted from the Kendlers' theory that mediators would do better than nonmediators. Confirmation of this prediction would support the verbal mediation hypothesis, tend to validate the mediation test and at the same time, incidentally, pose a problem for the dimensional orienting response model, which does not postulate a *verbal* mediating process; contrariwise, failure to confirm such a prediction would cast doubt on the validity of the mediation test, disparage somewhat the verbal mediation hypothesis and tend to support the dimensional orienting response model.

¹ This paper is based on the author's doctoral dissertation, written under the direction of N. A. Fattu. The pilot study for the present experiment was undertaken by the author in collaboration with John Murphy as a project for the Institute of Educational Research. Thanks are extended to Mr. Murphy for building the apparatus and assisting in the conduct of the pilot study and to Daphne Marlatt, who served as the second experimenter.

² Now at the University of Illinois.

To test the above prediction children identified by the mediation test as mediators and nonmediators were given an Osler-type concept-attainment task (Osler & Powell, 1960). Thus, only when the subject (*S*) attaches a common label to all members of the set of positive stimuli and motor-responds to the verbal *S*^D so produced (i.e., mediates) is he likely to solve the task in the allotted number of trials. It was expected from the verbal mediation hypothesis that mediators would do better than nonmediators on the concept-attainment task (Prediction 1).

In addition to mediation test category, two other variables were employed to form a $2 \times 2 \times 3$ factorial experiment (4 *Ss* per cell in the mediator groups; 6 *Ss* per cell in the non-mediator groups). Verbalization (overt condition versus nonovert condition) was chosen as one of these variables for two reasons: (a) To check, under a concept-attainment arrangement, a previous finding that overt verbalization facilitates performance in a discrimination learning situation (Weir & Stevenson, 1959). It was expected that the overt verbalization condition would be superior to the nonovert verbalization condition (Prediction 2). (b) To determine the differential effect (if any) of overt verbalization on the concept-attainment processes of mediators and nonmediators. Kendler (1963) has suggested that nonmediators do not mediate simply because they do not verbalize during the discrimination task. It was expected, therefore, that overt verbalization would facilitate the performance of nonmediators more than it would that of mediators (Prediction 3).

In introducing the third factor of the experiment—pretraining—an attempt was made to assess the transfer effect of a preestablished mediated

connection on concept attainment, as well as the differential effect of this connection in mediators and nonmediators. In Pretraining Condition 1 (PTC 1) the response-produced verbal cue, "bird," was (hypothetically) conditioned to an overt motor response of button pushing by means of a simple discrimination problem involving the concept-attainment task apparatus (cf. Jeffrey, 1953). Because of this connection, it was thought, any stimulus in the concept-attainment task (given immediately after) which elicited the verbalization "bird" should also elicit a button-push response as well. Since the concept to be attained was "bird," positive transfer was expected from PTC 1.

Pretraining Condition 2, which served as a specific control for PTC 1, was the same as PTC 1, except that "shoe" rather than "bird" was conditioned to the button-push response. As no shoe-pictures appeared as stimuli in the concept-attainment task, transfer from this connection per se was expected to be neither positive nor negative. It was predicted, therefore, that PTC 1 would be superior to PTC 2 (Prediction 4). Pretraining Condition 3 was simply a control for assessing all remaining transfer effects involved in the pretraining process itself, including warm-up, and consisted of no pretraining at all. All other PTC, it was expected, would be superior to PTC 3 (Prediction 5).

METHOD

Apparatus

Since the Kendler, Kendler, and Lennard mediation-test apparatus closely resembles that used in the present concept-attainment task, their equipment was modified to avoid unwanted transfer effects. The present mediation-test apparatus incorporates all essential features of the Kendler, Kendler, and Lennard equipment but was modeled after that used by Kendler and Kendler (1959). As the same theoretical principles are alleged to hold for both apparatus, this change was

not thought to seriously alter the character of the test. Apparatus for the mediation test consisted of four opaque tumblers and a small gray-enamelled turntable. The turntable was made of a 17-inch (diameter) circular board across the middle of which was placed a 10 × 17-inch plywood screen. Black rubber grommets 6 inches apart held the marble rewards which were baited under the tumblers between trials. Two of the tumblers were flat black and two were flat white. One of each color was 4.2 inches tall, 1.9 inch wide at the top and 2.8 inches wide at the base; the other two were 2.1 inches tall, 1.9 inch wide at the top and 2.5 inches wide at the base.

Apparatus for the pretraining and concept-attainment tasks were the same; only the stimuli were different. The gray-enamelled wooden apparatus consisted of a 30 × 24-inch panel in which were centered two windows (2.4 inches in diameter) 7 inches apart. Side panels (24 × 12 inches) supported the front panel while hiding the first experimenter (*E-1*), who operated the apparatus from the rear. All edges and windows were bordered in blue. Below each window (1.5 inch) was a red response button. Behind the panel, a light signaled *S*'s responses and enabled *E-1* to administer reinforcement appropriately. Marble rewards were delivered through a hole midway between the two buttons, 5 inches beneath them, and were caught in a plastic cup 1 inch below. Delivery of the marble, as well as presentation of the stimuli, was manually controlled. Between trials, *E-2* (who alone interacted with *S* in the session) lowered a green shade to cover the windows.

Stimuli for pretraining were a metallic blue equilateral triangle (1 inch sides), a red square (1.5 inch sides), and a yellow circle (1 inch diameter). All figures were made of construction paper and were glued to white backgrounds.

The concept-attainment task stimuli consisted of a pack of 20 different cards, on each of which were pasted two approximately 2-inch multicolored pictures (seals) spaced to fit the windows of the concept-attainment task apparatus. Each card presented pictures of a bird and either an animal or flower—the bird seals appearing equally often on the right and left sides of the cards in a prearranged nonsystematic sequence.

Small toys were given as prize incentives after the mediation test as well as after the pretraining and concept-attainment tasks.

Overview of Procedure

Table 1 shows the design of the experiment and order of procedure. A detailed

description of the experimental activities is presented in the next three sections.

Mediation Test Procedure

All *Ss* were administered the mediation test 2 to 3 days before the concept-attainment task. Procedure for administration conformed strictly to that prescribed by Kendler, Kendler, and Learned with the following exceptions: (a) Instructions were modified to make them apply better to the present apparatus and were slightly shortened. (b) Cut-off for Series 1 was placed at 100 trials rather than at 150 trials. Two *Ss* were thus eliminated (one trained on size, the other on brightness). The full procedure is described in Kendler, Kendler, and Learned (1962). Below, however, is a brief description.

The test was divided into three parts. In Series 1 the stimulus pairs (tall-white, short-black) and (tall-black, short-white) were presented one pair per trial in a prearranged sequence, and, for any given *S*, one of the four cues (i.e., black, white, tall, or short) was 100% reinforced by means of marbles baited under the positive stimuli. [Prior to Series 1, the positive cue for that series, as well as the stimulus pair to be used in Series 2, was determined randomly for each *S* with the restriction that each cue should be positive for an equal number of *Ss* during Series 1 and that the two stimulus pairs (i.e., tall-white, short-black, and tall-black, short-white) should be used with equal frequency as the Series 2 discriminanda.] Although the position of the positive stimulus normally alternated over trials from left to right throughout the series according to a prearranged pattern, stimuli remained constant over trials whenever *S* made an error. All *Ss* worked to a criterion of 9 out of 10 consecutive correct responses, at which point they were immediately shifted to Series 2.

In Series 2 only one of the stimulus pairs used in Series 1 was presented and the previously negative stimulus now became positive. Following *S*'s achievement of another 9 out of 10 criterion on this reversed discrimination, Series 3 commenced.

In Series 3 the pair of stimuli not used in Series 2 was reintroduced and alternated with the Series 2 pair in the same sequence as was used in Series 1. (Series 3, however, had a fixed length of 20 trials, each pair appearing 10 times.) Reinforcement of the stimuli used in Series 2 continued as in Series 2; both new stimuli (test pair) were baited, however, in order to avoid differential reinforcement. The Series 2 stimuli thus

TABLE 1
EXPERIMENTAL DESIGN AND ORDER OF PROCEDURE

Group	Session 1	Session 2	
		Pretraining	Concept-attainment task
Overt-verbalization condition			
PTC 1 M N-M	Mediation test	Learn S''bird''-R _{push}	Label stimulus before responding
PTC 2 M N-M	Mediation test	Learn S''shoe''-R _{push}	Label stimulus before responding
PTC 3 M N-M	Mediation test	No pretraining	Label stimulus before responding
Nonovert-verbalization condition			
PTC 1 M N-M	Mediation test	Learn S''bird''-R _{push}	Respond without speaking
PTC 2 M N-M	Mediation test	Learn S''shoe''-R _{push}	Respond without speaking
PTC 3 M N-M	Mediation test	No pretraining	Respond without speaking

Note.—M = mediator; N-M = nonmediator; PTC = pretraining condition.

served to keep *S* responding to the same cue as that to which he responded in Series 2. His responses to the test pair disclosed this cue (e.g., black). On its basis *S* was considered to have made either a reversal shift (8 of 10 choices of the test stimulus which was incorrect in Series 1; such a response pattern indicates that *S* had learned to respond in Series 2 to the cue which was negative in Series 1), a nonreversal shift (8 of 10 choices of the test stimulus which was correct in Series 1; such a pattern indicates that *S* had learned to respond in Series 2 to a cue within the dimension—e.g., size—which was *irrelevant* in Series 1), or an inconsistent shift (less than 8 of 10 choices of either member of the test pair). Following the procedure of Kendler, Kendler, and Learnard, *Ss* making reversal shifts were classified as mediators, and the remaining *Ss* were classified as nonmediators.

After Series 3, *S* was shown the stimuli of Series 2 and asked "Which one is the

winning cup?" and upon no verbal response, "What does it look like?" A third query, "How do you know?" used by Kendler, Kendler, and Learnard was inadvertently left out. The child was then given a prize and cautioned against speaking of the proceedings.

Pretraining Procedure

On the second or third day after administration of the mediation test, each *S* who received pretraining (i.e., each *S* in PTCs 1 or 2) was brought to the same room as before and was immediately seated faced away from the concept-attainment task apparatus and taught the "names" of the three pretraining stimulus forms. The *Ss* in PTC 1 learned to name the triangle form, "bird," and PTC 2 *Ss* learned to name the same form, "shoe." Both groups learned to name the square form, "house," and the circle form, "cookie." A typical anticipation paired-associates procedure was used to

teach the names of the forms and the criterion of learning was 15 consecutive correct trials.

Following this preliminary portion of the pretraining period *Ss* were given a two-choice simultaneous discrimination involving the three forms. The discrimination was presented by means of the concept-attainment task apparatus, the forms appearing in the two windows of the apparatus and *Ss* responding to the buttons directly under the forms. The circle and square forms alternated over trials as the negative stimulus, and the triangle form, which appeared in the left and right windows according to a Gellerman series, was always the positive stimulus. As in the mediation test, however, stimuli changed between trials only after a correct response. Correct responses were reinforced by means of marble rewards.

All *Ss* were required to correctly name whichever stimulus they chose before responding to it and were prompted when unable to do so. In addition *Ss* were made to repeat their choices whenever they responded before verbalizing. This additional trial was not counted, however. It was expected that this procedure would serve to condition the overt response of button pushing to the stimulus "bird" in the case of PTC 1 and "shoe" in the case of PTC 2.

Following achievement of criterion (9 out of 10 consecutive correct responses) *Ss* in PTCs 1 and 2 were given a prize and were immediately introduced to the concept-attainment task. Those in PTC 3 (i.e., the no pretraining condition) began the concept-attainment task immediately upon entering the room.

Concept-Attainment Task Procedure

Procedure for the concept-attainment task was similar to that of the pretraining discrimination in that stimuli appeared in the windows of the apparatus and *Ss* were required to respond to the buttons directly under the windows. In the concept-attainment task, however, a picture of a bird appeared in one window on each trial and a picture of some living thing other than a bird appeared in the other. Moreover, rather than the same positive stimulus and one of two negative stimuli appearing on each trial, as was the case in pretraining, within each block of 20 trials a different positive stimulus (bird picture) and negative stimulus (nonbird picture) appeared on each trial (i.e., there were pictures of 20 different birds and 20 different nonbirds and these pictures were repeated every 20 trials). As in pretraining, the posi-

tive stimulus appeared equally often in the left and right windows of the concept-attainment task apparatus.

Prior to the task all *Ss* were read common instructions regarding general task procedure. In addition, *Ss* in the overt verbalization condition were instructed to label the picture to which they wished to respond before responding. (Unlike in pretraining, however, *Ss* were free to apply any label to the stimulus that they wished.) The *Ss* in the nonovert verbalization condition were instructed to respond without speaking. All *Ss* attempted to attain the concept "bird," and attainment of the concept was considered to have occurred when *S* responded to a bird picture on 9 of 10 consecutive trials (criterion).

In Trials 1-3, which were not counted, *E* corrected any mistakes in *S*'s procedure. Every tenth trial *E* said, "Never push the losing button if you can help it; try to win a marble every time." Any *S* in the overt verbalization condition who failed to name a picture before responding to it was forced to name the picture and respond to it again immediately. This additional trial was not counted, however. At criterion or 100 trials the problem was stopped; *S* was then given his prize and cautioned not to tell anyone how to play the "game." This caution terminated the session.

Subjects

Sixty children (26 boys and 34 girls) drawn from three first-grade classes at the University Elementary School formed the sample. In addition to these children, six other *Ss* participated whose data were not reported. Two of these *Ss* failed the mediation test and were eliminated for this reason; in addition four *Ss* were randomly eliminated to obtain proportionality. Of the 60 children for whom data are reported, 24 were mediators and 36 were nonmediators. Within these mediation categories *Ss* were randomly assigned to either the overt or nonovert verbalization conditions and to one of three PTCs (four *Ss* per cell in the mediator group and six *Ss* per cell in the nonmediator group).

RESULTS AND DISCUSSION

Concept-Attainment Task

No significant differences between mediators and nonmediators nor between PTCs 1 and 2 were found in the pretraining data. Table 2 presents

TABLE 2
MEANS AND VARIANCES OF TRIALS TO
CRITERION FOR TWELVE GROUPS ON
THE CONCEPT-ATTAINMENT TASK

	Mediators		Nonmediators	
	Mean	Variance	Mean	Variance
Overt-verbalization condition				
PTC 1	2.0	6	29.6	1100
PTC 2	22.0	1206	23.6	600
PTC 3	63.0	953	56.7	542
Nonovert-verbalization condition				
PTC 1	63.8	1694	58.0	1539
PTC 2	65.8	403	89.1	403
PTC 3	87.5	469	66.6	742
Total	50.7	1930	53.6	1347

Note.— $N = 4$ for the mediators; $N = 6$ for the non-mediators.

the mean trials to criterion and variances of 12 groups on the concept-attainment task. A Hartley test indicated extreme nonhomogeneity of variance ($F_{\max(12,5)} = 276.5, p < .001$). Since an analysis of variance on these data revealed neither a main effect for mediation test category ($F < 1$) nor any significant interaction involving this variable, rather than risk losing a meaningful evaluation of the Verbalization $\times$ Pretraining interaction by transforming the scores, it was elected to collapse over the mediation test variable in order to eliminate the

nonhomogeneity in all subsequently reported analyses. Table 3 shows the new means and variances resulting from the collapse.

An $F_{\max}$ test on the new variances was not significant ($F_{\max(6,9)} = 2.25, p > .10$). The results of an analysis of variance on these data indicated that both verbalization ($F(1,54) = 20.3, p < .001$) and pretraining ($F(2,54) = 3.59, p < .05$) produced significant effects. The interaction was not significant ($F(2,54) = 2.01, p > .10$).

Since Predictions 4 and 5 deal specifically with the relative ordering of the PTC means, individual comparisons among the three PTCs were conducted. In line with Prediction 5, PTC 1 differed significantly from PTC 3 ($F(1,54) = 7.36, p < .01$); contrary to the same prediction, PTC 2 was not generally superior to PTC 3 ($F(1,54) = 2.30, p > .05$).

Contrary to Prediction 4, PTC 1 was not significantly different from PTC 2 ($F = 1.33$). The failure of this prediction to be fulfilled is somewhat anomalous in that it was strongly supported by a pilot study, particularly in overt-verbalization condition ($p < .001$). Because there were age differences in the pilot-study and parent-study Ss (pilot-study Ss were nursery schoolers), as well as changes in the concept-attainment task procedures

TABLE 3
MEANS AND VARIANCES OF TRIALS TO CRITERION FOR SIX GROUPS ON THE
CONCEPT-ATTAINMENT TASK

	PTC 1		PTC 2		PTC 3		Total	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Overt-verbalization condition	18.6	845	23.0	843	59.2	716	33.6	1171
Nonovert-verbalization condition	60.3	1607	79.8	1041	75.0	737	71.7	1239
Total	39.5	1748	51.4	1841	67.1	830		

Note.—All cells had $N = 10$.

used, PTCs 1 and 2 (overt verbalization condition) were replicated with 38 Ss from the University Nursery School (median age = 54 months; range 50–61 months) to isolate the effect of age. Again no difference was found between PTC 1 ($\bar{X} = 64.3$) and PTC 2 ($\bar{X} = 60.0$), $t < 1$. Thus, the present results provide no support for the hypothesis that the $s_{\text{bird}}-R_{\text{push}}$ connection theoretically involved in concept attainment is in any way strengthened through the pretraining procedure. Why this hypothesis was supported in the pilot study is not clear, although the pilot and present study differed in a large number of small details, any one of which might have been responsible for the discrepant results.

Though the data clearly show that overt verbalization facilitates concept attainment (and hence support Prediction 2), why it does is open to speculation. Weir and Stevenson's (1959) explanation is that verbalization requires *S* to orient to the relevant stimulus and thus enhances the likelihood of his sampling the correct cues, making the correct hypothesis, etc. This suggests, however, that overt verbalization condition superiority would be greatest in PTC 3, since both PTCs 1 and 2 also orient *S* to the appropriate stimuli. But Table 1 shows that the effect of verbalization is very much weaker in PTC 3 than in either of the other two conditions. Thus, some doubt is cast on this explanation.

Another hypothesis, suggested at the beginning of this paper, is that in the overt-verbalization condition a number of Ss are forced to label the stimuli who would not have done so otherwise. Since labels facilitate performance, the force of this explanation is apparent. On the other hand, it may be that overt verbalization simply increases the salience and discriminability of the verbal cue needed for

TABLE 4
FREQUENCY OF SUBJECTS FALLING IN THREE CATEGORIES AS A FUNCTION OF THE DIMENSION RELEVANT IN SERIES 1 OF THE MEDIATION TEST

Third series choice category	Dimension relevant	
	Size	Brightness
Reversal	19	5
Inconsistent	7	10
Nonreversal	5	14

concept attainment rather than actually generates it. Both interpretations account for the Weir and Stevenson data as well as the present results. Further research is needed, it would seem, to choose between them.

Both predictions involving the mediation test variable, it was previously pointed out, were unconfirmed by the exploratory $2 \times 3 \times 2$ analysis of variance. Thus, even though accepting the null hypothesis is a questionable tactic, it would appear that the Kendlers' mediation test is invalid for predicting the present concept-attainment task. Either of two tentative conclusions may be drawn from this inference: (a) The differential propensity to use verbal mediators measured by the mediation test is limited to simple tasks similar to the test itself. (b) Contrary to the verbal mediation hypothesis (and in line with the dimensional orienting response model) shift preference and ease of shift in concept-reversal tasks are not always functions of the presence or absence of *verbal* mediation. The mediation test data, as will now be shown, are rendered more intelligible by the latter alternative.

According to the verbal mediation hypothesis, whether an *S* chooses to make a reversal or nonreversal shift in the mediation test (i.e., responds as a mediator or nonmediator) depends on whether or not his behavior during

the test is guided by implicit verbal responses. Kendler (1964) has found, however, that choice of shift is determined to some extent at least by dimensional salience, and in the present mediation test dimensional salience exercised a considerable influence on this variable ($\chi^2_{(2)} = 12.89$; see Table 4). This fact would seem to render nugatory the contention that shift choice (in the present mediation test at least) is determined by the presence or absence of verbal mediators. It should be pointed out, however, that such a fact does not *prove* that verbal processes have little or no influence on shift behavior. Indeed, Ss may respond primarily to the prepotent dimension (here, size) because dimensional orientation is verbally controlled (as the verbal mediation hypothesis maintains) and size rather than brightness labels are dominant in the 6-year-old's verbal hierarchy. Interestingly enough, however, while more Ss responded to size than to brightness cues in Series 3 of the mediation test (size = 33, brightness = 10; $p < .001$, binomial test), the majority of Ss who verbalized a unique dimension labeled the stimuli with brightness words (brightness = 31, size = 12; $p < .01$, binomial test). Thus, motor choice, far from being *guided by*, seems actually to be *contrary* to the particular psychological processes determining verbalization. Since 40% of Ss were allegedly "mediators" this finding is difficult to reconcile with the Kendlers' theory. In consideration of this fact and the lack of relation between the mediation test and concept-attainment task found earlier, the present data would appear

to be more in line with the dimensional orienting response model of Zeaman and House (1963) than the verbal mediation hypothesis of Kendler and Kendler (1962).

REFERENCES

- HOUSE, B. J., & ZEAMAN, D. Reversal and nonreversal shifts in discrimination learning of retardates. *Journal of Experimental Psychology*, 1962, **63**, 444-451.
- JEFFREY, W. E. The effects of verbal and nonverbal responses in mediating an instrumental act. *Journal of Experimental Psychology*, 1953, **45**, 327-333.
- KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1-16.
- KENDLER, T. S. Development of mediating responses in children. *Monographs of the Society for Research in Child Development*, 1963, **28** (Whole No. 6), 33-51.
- KENDLER, T. S. Verbalization and optional reversal shifts among kindergarten children. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 428-436.
- KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology*, 1959, **58**, 56-60.
- KENDLER, T. S., KENDLER, H. H., & LEARNARD, B. Mediated responses to size and brightness as a function of age. *American Journal of Psychology*, 1962, **75**, 571-586.
- OSLER, S. F., & POWELL, M. G. Apparatus for the study of discrimination and concept formation. *American Journal of Psychology*, 1960, **73**, 627-629.
- SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427-449.
- WEIR, M. W., & STEVENSON, H. W. The effect of verbalization in children's learning as a function of chronological age. *Child Development*, 1959, **30**, 173-178.
- ZEAMAN, D., & HOUSE, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963, Pp. 159-223.

(Received February 18, 1966)

RELATIONSHIP OF SELECTED COGNITIVE ABILITIES TO CONCEPT ATTAINMENT AND INFORMATION PROCESSING¹

ELMER A. LEMKE
Illinois State University

HERBERT J. KLAUSMEIER
University of Wisconsin
AND CHESTER W. HARRIS
University of Wisconsin

Scores from 16 tests, 2 for each of 8 abilities (General Reasoning, Verbal Comprehension, Induction, Deduction, Spatial Scanning, Perceptual Speed, Rote and Span Memory), and 18 scores from concept-attainment and information-processing tasks were obtained from each of 94 female Ss enrolled in educational psychology at the University of Wisconsin. The 34 task and ability variables were inter-correlated, then factored using Alpha factor analysis. The 12 Alpha factors were rotated to an oblique solution according to the Harris-Kaiser criterion. 3 abilities (General Reasoning, Induction, and Verbal Comprehension) were found to be related to 3 concept-attainment and information-processing factors. The concept-attainment and information-processing tasks were seen to be relatively distinct rather than unitary activities.

In the scientific study of human learning, variables have been classified as input, intervening, and output. Although this classification is applicable to the study of concept attainment, a further breakdown of these categories is more typical of laboratory experiments in concept attainment completed thus far: (a) stimulus, referring to variables associated with the material in which the concepts are embedded; (b) instructions, referring to information presented to the subjects (Ss) in oral or written form concerning the task, procedures, and the like; (c) organismic variables of a cognitive nature, referring to the abilities, traits, etc., inferred from the test performance of the Ss; (d) organismic variables of a motivational nature, dealing with incentives, reinforcements, sets, etc.; and (e) response. While the goal of a compre-

hensive program of research² is to ascertain functional relationships among variables in all of the categories, the present study focuses primarily on organismic variables.

Such a program of research differs from that of Hovland (1952) and Hunt (1962), who have treated concept attainment as information processing. Their approach has been to treat concept learning and information processing as synonymous. Testing this line of thinking, Tagatz (1963) found only a small positive correlation, .22, between information processing and concept attainment. Tagatz defined information processing as the ability of an S to specify the inclusion, exclusion, or indeterminance of a card to membership in a group of cards which are specified as belonging or not belonging to a concept. He defined concept attainment as the ability to utilize information from positive and/or negative instances in determining a

¹The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

This article is based on the doctoral dissertation of the senior author.

²Concept learning and related cognitive abilities is the main focus of research of the R & D Center for Learning at the University of Wisconsin.

concept. (See pp. 29-30 for a more detailed description of concept attainment and information processing as defined in this study.)

Despite the substantial amount of research on concept attainment, the relationship of organismic variables to concept attainment has been given little attention. This investigation is specifically concerned with identifying those factors or abilities that are highly related to concept attainment and information processing. After a review of the hypothesized factors of intellect and tests intended to measure them, eight factors were selected for study based on their presumed relevance or lack of relevance to concept attainment and information-processing tasks.

A brief description of these eight factors, based on French, Ekstrom, and Price (1963), and an indication of their presumed relationship to concept attainment will clarify the rationale of the present study.

Rote Memory is defined as the ability to retain bits of unrelated material. When an *S* first encounters a large amount of stimulus material of the type used in this study, he may perceive it as being unrelated (although it is actually highly related). To attain a concept wherein information is tested successively, as in the present study, an *S* must be able to retain the information. Some type of memory appears to be required for efficient concept attainment as defined in Task 1 of this study. (Task 1 is described more fully in the next section.)

Span Memory involves the ability to recall perfectly for immediate production a series of items after only one presentation. Although the mode of presentation was simultaneous in the concept-attainment tasks, *S*'s identification of instances as exemplars or

nonexemplars was sequential in Task 1. The instances' sequential identification by number was seen as a basis for including the Span Memory factor.

Perceptual Speed involves speed in finding figures, making comparisons, and carrying out other very simple tasks involving visual perception. Both concept-attainment tasks and the information-processing tasks of the present study presumably require the ability to make comparisons of figural material. The *S* must discriminate among visual stimuli and make comparisons in order to secure essential information.

General Reasoning is the ability to solve a broad range of problems that require production of a generally accepted correct solution, including those of a mathematical nature. Although the stimulus material in the present study is not mathematical, attaining a concept presumably requires the ability to compare information and to arrive at a correct solution.

Deduction involves the ability to reason from stated premises to their necessary conclusions. In the present study, the information-processing tasks might presumably have required this ability in that propositions of a similar type comprised the items of the information-processing task: If the focus card is a member of the group, and the second stimulus card is also, does the third card definitely belong to the group, definitely not belong to the group, or can its membership not be determined?

Induction probably involves several abilities associated with the finding of general concepts that will fit sets of data, the forming and trying out of hypotheses. The concept-attainment tasks presumably involved these abilities directly; to a lesser extent the in-

formation-processing tasks presumably did also.

Spatial Scanning requires the ability to explore visually a wide or complicated spatial field. Finding one's way through a paper maze is a test of this ability. A planning ability may also be involved. If this ability is related to any task in the present study, it should presumably be concept-attainment Task 1, not the other performance tasks.

Verbal Comprehension is the ability to understand the English language. The importance of the factor in both the British factor hierarchy and the Thurstone studies suggested its inclusion in the investigation.

The present study, then, was designed to clarify relationships among cognitive abilities, information processing, and concept attainment. A factor analysis, to an oblique criterion, of the ability (cognitive variables) and task (information-processing and concept-attainment criterion variables) scores provided the model for the study. The matrix of intercorrelations of ability and task factors resulting from the oblique factor rotation provided the desired relationships.

METHOD

Tasks

Two concept-attainment tasks and an information-processing task were utilized in the experiment. The first task, a concept-attainment task, is described by Harris (1963):

The experimenter (E) chooses the concept to be attained. He then points out a card on a display of 64 cards that belongs to this group; this will be called the focus-card or (F). The subject (S) is then directed to choose one card at a time from the display; for each choice E responds "yes" if the chosen card belongs to the group, "no" if it does not. S is told initially to specify the defining characteristics of the card group (the concept) when-

ever he believes he has solved the problem; if his "hypothesis" is correct, the game terminates; if not, he is told "no" and asked to continue the card choice. For such play, S's card choices in the order in which he makes them, and the hypotheses he offers in their order and in relation to his card choices, or solution, may be recorded.

In the second concept-attainment task S was confronted with a small number of cards presented simultaneously. Unless redundancy was introduced, only the minimum number of cards necessary to attain the concept was presented. Thus, S was presented a small display and was requested to state the concept into which all the cards could be classified. The information-processing task is described by Tagatz (1963):

That part of the experiment dealing with information processing consisted of the S responding to 60 items. The first 30 were of the type in which one card, either an exemplar or a non-exemplar, was presented in addition to the exemplar focus card. The task was to specify the inclusion, exclusion, or indeterminateness of a third card to membership in a group of cards exemplifying the concept. The problems in which exemplars were presented numbered 15 and those presenting non-exemplars also 15. Of these 15 exemplar items, ten cards whose membership was to be determined were definitely exemplars of the same concepts as the focus card. The membership of the remaining 5 test cards could not be determined. In the 15 problems presenting a focus card and a non-exemplar, 10 test cards were definitely not exemplars and 5 were again not determinable. Thus these 30 items could be scored on the basis of test membership exemplar, non-exemplar, and indeterminate.

The second set of 30 items was constructed, using the same focus card and test cards as in the first subtest. The information presented in addition to the exemplar focus card consisted of two cards instead of one as was the case with the first subject. One of the two cards for each problem was an additional exemplar; the other was the same in kind as its counterpart in the first 30 items. The answers to the 30 items of the second subtest were identical to those in the first subtest. Thus, the two subtests were the same except that the information presented about the test

TABLE 1
SUMMARY DESCRIPTION OF EIGHT
INFORMATION-PROCESSING SUBTESTS

Subtest	First card	Second card	Decision card	No. of cards
1	yes	—	Inclusion	10
2	yes	—	Indefinite	5
3	no	—	Exclusion	10
4	no	—	Indefinite	5
5	yes	yes	Inclusion	10
6	yes	yes	Indefinite	5
7	yes	no	Exclusion	10
8	yes	no	Indefinite	5

card in the second subtest included the additional complexity of one card.

Table 1 gives the design of the arrangement of exemplar and/or nonexemplar instances and number of cards or items per test. As shown in Table 1, this information-processing task resulted in eight subscores of information processing based on the type of information contained in the stimulus material.

Task 1, in which *S* selected instances from a total display, had these measures of efficiency: time required to attain the concept, an index of manifested information, and total number of cards chosen prior to attaining the concept. Task 2, wherein a minimally sufficient set of information was presented to *S*, had time to attain the concept as the index of efficiency. In the information-processing task, scores on each of eight subtests were performance criteria. The index of manifest information in Task 1 was defined as amount of information manifested in the first hypothesis, or statement of the concept, from that potentially obtained. Thus, if five bits were potentially obtained but only three were manifested in the hypothesis, the index was .60. The *Ss* were asked for a concept after their sixth card choice if one had not been previously offered.

Stimulus materials used in this investigation were patterned by Byers (1961) after the Wisconsin Card Sorting Task. The concept-attainment display consisted of an ordered arrangement of attributes, by rows and columns, which formed a 8×8 array of 64 cards. On every card six attributes were presented by one of two defining characteristics. These six attributes and their dichotomous defining characteristics were:

1. Border number: one—two
2. Border continuity: solid—broken

3. Figure number: one—two
4. Figure size: large—small
5. Figure color: red—green
6. Figure shape: circle—ellipse

Each card was different from all other cards on the display in at least one of the dichotomous defining characteristics.

Tests

A battery of 16 ability tests, two for each of eight factors previously mentioned, was administered to all *Ss*. The tests were from the Reference Kit for Cognitive Factors (French et al., 1963). Table 2 presents the factor, number of items, and reliability coefficients for each test. Reliability coefficients for variables measuring Perceptual Speed and Spatial Scanning factors are not reported because of the spuriously high coefficients resulting from the obvious speeding of these tests.

Subjects

The *Ss* for the study were 94 graduate and undergraduate females from two educational psychology classes. All *Ss* participated in four sessions of group testing and a fifth individualized concept-attainment session. All *Ss* were in the age range of 20–35 and had a median age of 21 years.

Procedure

Each *S* attained a set of two selection concepts and a set of four presented concepts. For the first two concepts, *S* selected instances from a display in which all the stimulus material was presented simultaneously. The concepts to be attained were of two and three relevant attributes, for example, small, red figures; two borders, small, red figures. *Ss* were randomly assigned to one of two sequences in order to control for the complexity effect of the concepts. The sequences were: two-attribute concept, three-attribute concept; or three-attribute concept, two-attribute concept.

After attaining the first two selection concepts, each *S* attained four concepts from a minimally sufficient set of simultaneously presented information. These four concepts were solved in a sequence with positive and negative instances varying one attribute from the focus card in the first two concepts. In the last two concepts to be attained, negative instances were varied one attribute, but only one positive instance was used to attain the minimally sufficient set of information. The sequence was comprised

TABLE 2

RELIABILITY ESTIMATES OF IDENTIFICATION VARIABLES FOR EIGHT HYPOTHESIZED FACTORS

Factor	Test	No. of Items	Procedure	r_{tt}
Rote Memory	Object-Number	15	Split-half	.68
	First & Last Names	30	Split-half	.77
Verbal	Vocabulary	36	Split-half	.86
	Advanced Vocabulary	36	Split-half	.81
Deduction	Logical Reasoning	40	Split-half	.72
	Nonsense Syllogisms	30	K. R. 20	.88
Span Memory	Auditory Number Span	24	K. R. 20	.62
	Auditory Letter Span	24	K. R. 20	.76
General Reasoning	Ship Destination	57	Split-half	.93
	Necessary Arithmetic Operations	15	Split-half	.74
Perceptual Speed	Finding A's	25		a
	Number Comparison	24		a
Induction	Locations	28	Split-half	.82
	Letter Sets	30	Split-half	.64
Spatial	Map Planning	20		a
	Maze Tracing Speed Test	24		a

a Traditional computational techniques not appropriate.

of concepts of two, three, two, and three relevant attributes.

The information-processing task was administered in a large group-testing session after the individually administered concept-attainment session. Stimulus materials, or instances of concepts, were presented on 3×3 slides. Each slide thus comprised an item and was subsequently scored as correct or incorrect.

RESULTS

Two scale-free models, Incomplete Image analysis (Harris, 1962), and Alpha Factor analysis (Kaiser & Caffrey, 1963), which rescale the reduced correlation matrix in the metric of the unique and common parts respectively, were employed as data reduction models. Only the Alpha model is reported in this discussion.

A Normal Varimax rotation (Kaiser, 1958) of the Incomplete Image factors yielded 23 factors, eight of which were uninterpretable. The same

orthogonal criterion applied to the Alpha factors yielded a derived matrix of 12 factors, all of which were interpretable. The 12 Alpha factors, when rotated to the Normal Varimax criterion, were found to include the following factors (Roman numerals in Table 3) identified by the experimental tests of cognitive abilities: (I) Verbal Comprehension, (IV) Rote Memory, (VI) Span Memory, (VII) Spatial Scanning, (X) Deduction, (XI) General Reasoning, and (XII) Induction. The Perceptual Speed factor was not identified. Thus seven of the eight ability factors which the 16 experimental tests were supposed to measure were in fact identified.

Five factors identified by the loadings of the task variables—three concept-attainment and two information-processing factors—were called (Table 3): (II) Information Processing (In-

TABLE 3
ROTATED FACTORS OF H^{-1} ($R-S^2$) H^{-1}
(OBLIQUE CRITERION)

Test	V. I	I.P. II	Sel. III	R.M. IV	Pres. VI	S.M. VI	S.S. VII	Sel. VIII	I.P. IX	D. X	G.R. XI	I. XII
1. Object-Number	-335	054	-062	689	162	064	066	-029	189	079	-052	062
2. First and Last Names	241	072	029	594	-163	-007	-101	-039	024	-006	-049	002
3. Vocabulary	796	035	049	038	038	072	-009	150	-038	056	009	021
4. Advanced Vocabulary	900	-112	010	000	-013	-001	037	-080	005	051	021	-142
5. Logical Reasoning	074	-077	043	042	061	-129	-013	-042	077	754	056	063
6. Nonsense Syllogisms	035	255	-431	-104	066	123	-122	-082	388	554	554	-136
7. Auditory Number Span	-026	-049	055	010	-009	839	068	-094	025	-039	113	-024
8. Auditory Letter Span	081	-061	012	072	019	750	084	052	016	-106	004	084
9. Ship Destination	-066	017	022	-212	030	015	213	-143	072	-068	711	094
10. Necessary Arithmetic Operations	005	-020	171	018	-127	025	-011	-298	166	157	866	-213
11. Finding A's	055	-112	-116	303	022	110	009	040	-307	-034	438	098
12. Number Comparisons Test	178	-040	093	221	031	-305	843	-112	-304	-084	074	046
13. Locations Test	-247	060	097	-060	-043	057	-071	172	-192	139	687	804
14. Letter Sets	-073	-034	036	074	037	019	-019	-083	-010	023	-027	809
15. Map Planning	016	080	093	-042	093	178	696	046	118	-020	-039	-078
16. Maze Tracing Speed Test	005	-117	-209	-054	-186	-085	484	250	-050	027	154	386
17. Selection 1, Time-to-Criterion	045	-099	-025	-035	314	-118	117	561	147	-110	253	-170
18. Selection 2, Time-to-Criterion	133	136	893	-059	268	180	073	189	-069	025	-030	-021
19. Selection 1, Manifest Index	-130	-155	414	006	-014	-074	-082	215	114	-072	159	012
20. Selection 2, Manifest Index	-106	-109	499	015	-349	-155	064	006	068	035	292	-066
21. Selection 1, Cards-to-Criterion	007	030	-050	038	-152	014	061	921	056	025	-150	-041
22. Selection 2, Cards-to-Criterion	134	381	478	-085	-168	322	-061	281	-179	288	-275	175
23. Presented Concept 1	238	034	256	-042	291	-198	-102	169	-105	-032	-006	200
24. Presented Concept 2	098	-020	118	033	483	-146	138	013	-016	009	162	-030
25. Presented Concept 3	092	019	-125	037	373	067	-104	138	-163	-075	357	-117
26. Presented Concept 4	-125	069	069	-034	712	013	102	141	117	102	-154	078
27. Information Processing, Yes-Inclusive	117	692	-001	144	-139	-088	048	-040	-068	-083	163	064
28. Information Processing, Yes-Indeterminate	078	112	-121	-006	113	-005	-054	014	457	028	-061	376
29. Information Processing, No-No Exclusive	-010	735	-101	735	041	-035	082	-086	048	-176	037	-076
30. Information Processing, No-Indeterminate	112	227	038	-161	-026	-028	-213	-120	191	-328	443	223
31. Information Processing, Yes, Yes-Inclusive	-026	176	072	176	-150	-108	151	128	512	-017	189	-191
32. Information Processing, Yes, Yes-Indeterminate	-027	-111	-022	173	095	059	049	062	654	042	-007	016
33. Information Processing, Yes, No-Indeterminate	198	-132	132	097	077	017	-305	116	811	-110	190	080
34. Information Processing, Yes, No-Exclusive	-182	726	110	063	107	-028	016	021	-004	087	-111	-040

clusion-Exclusion), (III) Selection (Concept 2), (V) Presented Concept, (VIII) Selection (Concept 1), (IX) Information Processing (Indeterminate).

To observe ability-task relationships, the Alpha factor matrix was rotated to an oblique solution using the Harris-Kaiser (1964) criterion. Since some variables were of complexity greater than one, the A'A proportional to L case was used. Table 3 represents the 34×12 oblique factor matrix.

Italic loadings of Table 3 provided the rationale for oblique factor descriptions. Only the Induction factor, deduced from the Letter Sets and Locations tests, presented some difficulty in identification. Thurstone's (1940) isolation of this factor from the Letter Grouping test facilitated the factor identification. (See also Goodman, 1943; Kettner, Guilford, & Christensen, 1959; Thurstone, 1941). The Spatial Scanning factor, previously isolated but unidentified by Thurstone and Thurstone (1941) and suggested by French et al. (1963) as representing a "planning" function, is also of interest. The strong involvement of this factor with tests from the reasoning domain suggests that there is, in fact, a strong convergent involvement in this type of activity.

The L matrix, the matrix of factor intercorrelations for the 12 task and ability factors, is presented in Table 4. Of 21 correlations among seven ability factors (Table 4), seven were positive and ranged between .238 and .579; Verbal Comprehension, General Reasoning, and Induction are involved in these seven correlations. Of 10 correlations among three concept-attainment and two information-processing tasks, five were positive and ranged between .213 and .431; the factor Information Processing—Indeterminate Inclusion, was involved in three of five positive correlations. Thus correla-

TABLE 4
MATRIX OF INTERCORRELATIONS OF 12 ALPHA FACTORS

Factor description	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1. Verbal Comprehension	1.000											
2. Information Processing, Inclusion-Exclusion		.073	.164	.092	.220	.016	-.002	.198	.317	.289	.425	.350
3. Selection I		1.000	.111	.141	.334	.172	.144	.171	.431	.085	.397	.259
4. Rote Memory			1.000	.105	.058	.098	.060	.415	.157	.116	.366	.353
5. Presented Concept				1.000	.033	.038	.194	-.011	.089	-.020	.165	.200
6. Span Memory					1.000	.168	.106	.166	.213	.145	.461	.338
7. Spatial Scanning						1.000	.047	.116	.173	.095	.128	.057
8. Selection II							1.000	.124	.025	.089	.305	.249
9. Information Processing, Indeterminate								1.000	.278	.167	.363	.229
10. Deduction									1.000	.150	.411	.313
11. General Reasoning										1.000	.238	.127
12. Induction											1.000	.579
												1.000

Note.—All decimal points eliminated for off-diagonal elements.

tions among variables within each of two sets of variables were of about the same magnitude and the proportion of the total was about the same.

Of 35 correlations between the set of seven ability variables and the set of five task variables, 12 were positive and ranged from .220 to .461. Of these 12 correlations, 10 involved General Reasoning and Induction; the other two, Verbal Comprehension. General Reasoning correlated positively with all five of the task factors, the range being from .363 to .461. Induction also correlated with all five task factors, the range of r 's being .229 to .353. Thus General Reasoning, Induction, and, to a lesser extent, Verbal Comprehension correlated substantially and consistently with concept attainment and information processing; the other four abilities—Rote Memory, Span Memory, Spatial Scanning, and Deduction—did not. Further, Information Processing—Indefinite Inclusion was the information-processing factor that correlated most consistently with other task factors and the cognitive factors, six of 11 r 's ranging between .213 and .431. Presented Concept was the concept-attainment factor to correlate most consistently with task and cognitive ability factors, five of 11 r 's ranging between .213 and .461.

DISCUSSION

Low correlations were found between a set of cognitive abilities, a set of concept-attainment factors, and a set of information-processing factors. However, several abilities—Verbal Comprehension, General Reasoning, and Induction—correlated consistently with the concept-attainment and information-processing factors.

The identification of three relatively distinct concept-attainment factors and two information-processing factors with only a few correlations of

modest size among these factors was not anticipated. Consider the three concept-attainment factors. Two of these—Selection Concept I and Selection Concept II—resulted from measures of Ss' attaining two concepts in sequence, under identical experimental conditions; that is, in one sitting with identical instructions, materials, etc. Why did two separate factors result? Ss first attained a two-attribute and then a three-attribute concept or first a three-attribute and then a two-attribute concept. They were not informed of this attribute change which affected the difficulty level of the concept. Apparently, the change in the number of attributes and the ordinal position affected performance in such a manner as to result in separate factors.

The third factor, Presented Concept, resulted from a very different experimental situation. Here the task was for Ss to attain four concepts of two and three attributes in sequence; however, only the minimum amount of information necessary to attain the concept was presented. Thus, attaining the concept under these conditions was distinctly different from that in which a large array was presented simultaneously and Ss selected cards successively.

Information processing as defined by eight subtests also yielded two factors, one being based on items of the type where the card definitely was or was not a member of the concept, the other being based on items where insufficient information was presented to determine inclusion in the concept. Identification of the two factors suggests that information processing is not a unitary ability, much the same as concept attainment is not. Further, information processing as defined in this study is not closely related to concept attainment.

The above considerations lead to a further question concerning the na-

ture of concept attainment and information processing, using stimulus material of the type included in the present study. How unique are tasks that result, for example, in three concept-attainment and two information-processing factors? How many factors would be found if various sequences of one-, two-, three-, four-, five-, and six-attribute concepts were used? These questions remain unanswered; however, the identification of the task factors suggests clearly that concept attainment and information processing are quite different, that these tasks appear to be quite different from the cognitive factors identified in this study.

REFERENCES

- BYERS, J. L. Strategies and learning set in concept attainment. Unpublished doctoral dissertation, University of Wisconsin, 1961.
- FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. *Manual for kit of reference tests for the cognitive factors*. Princeton, N. J.: Educational Testing Service, 1963.
- GOODMAN, C. H. A factorial analysis of Thurstone's sixteen primary mental abilities tests. *Psychometrika*, 1943, **8**, 141-151.
- HARRIS, C. W. Some Rao-Guttman relationships. *Psychometrika*, 1962, **27**, 247-263.
- HARRIS, C. W. The problem of strategy identification. Paper read at American Educational Research Association, Chicago, February 1963. (Mimeo)
- HARRIS, C. W., & KAISER, H. F. Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, 1964, **29**, 347-362.
- HOVLAND, C. I. A communication analysis of concept learning. *Psychological Review*, 1952, **59**, 461-472.
- HUNT, E. B. *Concept learning: An information processing problem*. New York: Wiley, 1962.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187-200.
- KAISER, H. F., & CAFFREY, J. Alpha factor analysis. *Psychometrika*, 1965, **30**, 1-14.
- KETTNER, N. W., GUILFORD, J. P., & CHRISTENSEN, J. P. A factor analytic study across the domains of reasoning, creativity, and evaluation. *Psychological Monographs*, 1959, **73**(9, Whole No. 479).
- TAGATZ, G. The effects of selected variables upon information processing and information utilization in concept attainment tasks. Unpublished doctoral dissertation, University of Wisconsin, 1963.
- THURSTONE, L. L. Primary mental abilities. *Psychometric Monographs*, 1940, No. 1.
- THURSTONE, L. L., & THURSTONE, T. G. Factorial studies of intelligence. *Psychometric Monographs*, 1941, No. 2.

(Received February 25, 1966)

EFFECTS OF SOCIAL CUES AND TASK COMPLEXITY IN CONCEPT IDENTIFICATION¹

AARON WOLFGANG²

Behavioral Science Laboratories, Veterans Administration Hospital, Oklahoma City

The 120 Ss participated in a factorial design which included 3 levels of task complexity and 3 conditions where Ss served individually or with a partner. In the free interaction (FI) condition 2 Ss were free to communicate and in the restricted interaction (RI) condition communication between Ss was limited. Ss in the FI condition outperformed individuals (only on the most complex concepts) and Ss in the RI condition across all levels of complexity. The assumption of the mathematical theory of CI that learning rate should decrease with increases in irrelevant information was confirmed in the individual and RI conditions but not in the FI condition.

Learning frequently occurs in groups where individuals exchange information, yet the most influential learning theories, including the more recent mathematical models (e.g., Estes, 1963; Restle, 1955), have postulates based solely on the behavior of the individual performing alone. Bourne and Restle (1959), in attempting to develop a comprehensive mathematical theory of concept identification (CI), have not as yet considered social cues and their influence on learning. In dealing with the effects of a wide variety of variables on CI, investigators have focused on the performance of individual subjects (Ss). Recently, Pishkin and Blanchard (1963), in extending the mathematical theory of CI to include social parameters, have established the value of social cues in a situation where a programmed stooge provided social cues but where communication between Ss was prohibited. Research dealing with the social aspects of CI where Ss are

free to discuss their hypotheses regarding the solution of problems varying in amount of irrelevant information or complexity has received little or no attention. The purpose of the present study was to compare learning rates of two-man groups free to communicate with that of single Ss and two-man groups where communication was restricted in solving CI problems of different complexities.

METHOD

Subjects

The Ss were 120 male volunteer students in elementary psychology courses, who were randomly divided into nine treatment conditions. The Ss in the two-man groups who acknowledged that they were friends were excluded from participating with each other.

Design

A 3×3 factorial design was used, which included three levels of complexity (1, 3, or 5 irrelevant bits of information) and three levels of interaction. The basic unit of measurable relevant and irrelevant information in information theory is a bit, which is defined as $\log_2 x$ where x is the number of equally probable stimulus events. The general rule is that every time the number of alternatives from which to make a decision is increased by a factor of 2, 1 bit of information is added. In the present experiment the stimuli to be classified were specified along different binary stimulus dimensions (e.g., size: large or small; position: top or bottom). There was always 1 bit of rele-

¹ This article is based on a doctoral dissertation submitted to the Graduate School of the University of Oklahoma. Special thanks to Vladimir Pishkin for valuable suggestions. This project was supported in part by the Veterans Administration Medical Research Program (8200 Funds).

² Also at the University of Oklahoma Medical Center.

vant information accompanied by 1, 3, or 5 irrelevant bits. Thus for problems with 1 irrelevant bit there were 4 stimuli to categorize; with 3 irrelevant bits, 16 possible stimuli; and with 5 irrelevant bits there were 64 stimuli from which to make a choice.

In each of 24 two-person groups in the free interaction condition (FI), Ss were free to discuss their hypotheses or ideas with each other but had to reach a mutually agreeable decision on each trial before proceeding to the next trial; in the restricted interaction condition (RI) no discussion was permitted between Ss in each of the 24 two-person groups, Ss simply stated and registered their individual decisions on one of the two response keys on each trial; and lastly, 24 Ss performed in the individual condition (I).

Task and Apparatus

In solving a two-choice CI problem, S's task was to categorize a series of geometric patterns flashed on the screen in accordance with a relevant binary stimulus dimension. For example, where the relevant dimension was form, S would press key A in response to a square, and key B in response to a triangle and be correct. If, however, S responded to the irrelevant dimensions such as color (red or green) which has a zero correlation with the correct response, then his responses would be correct only at the chance level. The relevant dimensions of the two problem types were form and number.

The standard CI apparatus was used (Pishkin, 1960) with several modifications—experimenter's (E's) panel board and a semi-soundproof cubicle—described below.

The E's panel board was electronically connected to S's panel and contained, in an identical manner, two response keys identified by the letters A and B and two feedback lights; but the panel boards differed in function. When S pressed a key E's panel light indicated S's choice of response; then E, using a planned program of feedback coordinated with the filmstrip programming, depressed a key which lit up one of the feedback lights on S's panel for approximately 1 second, indicating to S the correctness of his response. The Esterline Angus 20-pen operations recorder was electronically connected to both E's and S's panel boards to record S's response and E's feedback.

To reduce background noise from the apparatus and surroundings, a wooden cubicle lined with soundproof tiles was constructed. The three-paneled cubicle with a top and

two sides was 63 inches high, 36 inches from front to back and 48 inches in width. It was roomy enough so that one or two Ss could be comfortably seated. The cubicle was arranged so that Ss could clearly view only the screen and panel board directly in front of them.

Procedure

At the start of each session, Ss in the social conditions were seated inside the soundproof cubicle in front of the screen, each having access to the response keys. The Ss were told about the nature of the task and the significance of the response keys and feedback lights. The instructions for individuals and Ss in two-person groups were essentially the same except that Ss in the FI group were told that they were to arrive at a single decision and only one decision could be registered per trial by either S; whereas Ss in the noninteracting control group were simply told to state their individual decisions verbally and then register them by depressing one of the two response keys on each trial.

The criterion to solution for Ss in all conditions was 16 consecutive correct responses. In the RI condition, criterion to solution occurred when each S independently made 16 consecutive correct responses. However, if criterion was not reached, then Ss were given a maximum of 192 trials.

RESULTS

Analysis of Error and Time Scores

Due to the marked heterogeneity (the ratio between the largest and smallest variance was over 100), a log transformation upon error and time scores was performed resulting in homogeneity of variance. An analysis of variance on log error scores disclosed significant main effects for levels of interaction ($F = 27.12$, $df = 2/63$, $p < .001$) and complexity ($F = 10.40$, $df = 2/63$, $p < .001$). Mean numbers of errors were 29.75, 13.5, and 2.29 in the RI, I, and FI conditions, respectively. Duncan's multiple-range test revealed that Ss in the FI condition were superior to those in the I ($df = 63$, $p < .005$) and RI conditions ($df = 63$, $p < .001$). The Ss in the I con-

dition made significantly fewer errors than those in the RI condition ($df = 63$, $p < .001$).

Figure 1 presents mean errors between interaction conditions at each level of complexity. To compare performance differences between interaction conditions at each level of complexity, t tests were computed. Although Ss in the FI condition made fewer errors than those in the I condition along all levels of complexity, significance was reached only at the highest level of complexity ($t = 3.70$, $df = 63$, $p < .001$). In contrast, Ss in the FI condition significantly outperformed those in the RI condition on problems with 1 ($t = 3.33$, $df = 63$), 3 ($t = 4.90$, $df = 63$), and 5 irrelevant bits ($t = 4.62$, $df = 63$) with $p < .01$. Individuals made significantly fewer errors than Ss in the RI condition on problems with 1 ($t = 2.85$, $df = 63$, $p < .01$) and 3 irrelevant bits ($t = 3.44$, $df = 63$, $p < .01$) but on problems with 5 irrelevant bits no significant differences were found ($t = .92$, $df = 63$, $p > .05$).

The significance of the complexity main effect showed that as the amount of irrelevant information increased mean errors progressively increased. The mean number of errors for prob-

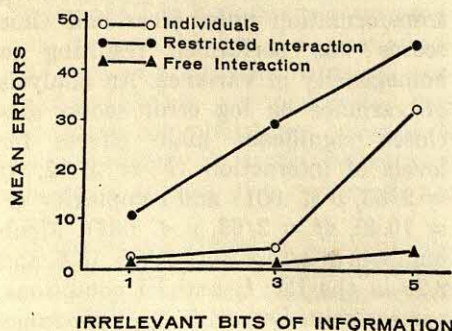


FIG. 1. Mean errors as a function of irrelevant bits of information. (The parameter is levels of interaction. Each point represents a mean based on an N of 8.)

lems containing 1, 3, or 5 irrelevant bits of information were 4.75, 12.12, and 28.66, respectively. An orthogonal polynomial analysis was performed for complexity and only the linear component reached significance ($F_{lin} = 19.99$, $df = 1/63$, $p < .001$).

An analysis of variance of mean log time to solution in minutes revealed significant main effects for levels of interaction ($F = 18.54$, $df = 2/63$, $p < .001$) and task complexity ($F = 15.12$, $df = 2/63$, $p < .001$). Subsequent analysis of levels of interaction with Duncan's multiple-range test indicated that there were insignificant differences in time to solution between Ss in the FI and I conditions ($df = 63$, $p > .05$). However, there were significant differences in performances between Ss in the FI and RI condition ($df = 63$, $p < .001$), with the Ss in the RI condition taking longer to reach solution. Mean times to solution in minutes for Ss were: FI-6.42, RI-17.50, and I-8.16. The significant main effect of complexity indicated that differences among mean times to solution were a function of the amount of irrelevant information. Mean times to solution in minutes for problems with 1, 3, and 5 bits of irrelevant information were 5.58, 9.75, and 16.79, respectively.

DISCUSSION

In the present study task complexity proved to be important in clarifying the generalization that cooperative groups are superior to individuals in learning. That FI groups outperformed individuals only on the most complex problems where information load, memory requirements, and number of alternative hypotheses were greatest could be attributed to several factors. First, opportunity for perseverating on an incorrect hypothesis is reduced in the FI group, since an S would have to

offer his partner some justification for his persistently incorrect responses whereas in the individual condition no such justification was necessary. Second, and probably more important, FI groups have more resources available and can draw on two memories for retaining and evaluating past information. The importance of memory in concept learning has been demonstrated by Cahill and Hovland (1960), Hunt (1961), Bourne, Goldstein, and Link (1964), and Pishkin and Wolfgang (1965). Recently, Laughlin (1965) presented evidence indicating that two-person cooperative groups showed better memory in concept attainment than individuals and made more efficient use of memory than individuals.

Judging from the relatively poor performance of the RI group, it is plausible that more irrelevant distracting cues were involved. The CI task can be described as one that requires sustained attention, concentration, reliance on past information feedback, and the ability to identify and classify relevant cues. The Ss in the RI condition might have had difficulty in concentrating on their own hypotheses and feedback when at the same time additional information was offered by their partners who were voicing and registering similar and contrasting responses. That learning was more retarded in the RI condition than in the individual condition was consistent with the results of a number of animal and human studies reported by Zajonc (1965).

In view of the slight deterioration in performance on problems of increasing complexity by Ss in the FI condition, it seems reasonable to expect that they would be able to solve problems of even greater complexity. At present the maximum number of irrelevant bits of information in two-

choice CI problems has been five. A logical extension of the present study would be to explore the limits of individual capacity to handle information of higher levels of complexity (e.g., 6 or 7 irrelevant bits) under social and individual conditions.

One of the basic postulates of the mathematical theory of CI is that learning rate should decrease with increases in irrelevant information (Bourne & Restle, 1959, p. 288). However, that assumption was confirmed in the I and RI conditions but not in the FI condition where Ss showed no decrease in learning rate from problems with 1 to 3 irrelevant bits. That learning was not progressively retarded with increased task complexity in the FI condition was inconsistent with the results of past investigators using individual Ss (e.g., Archer, Bourne, & Brown, 1955; Bourne, 1957; Bourne & Haygood, 1959; Pishkin, 1960; Wolfgang, Pishkin, & Lundy, 1962). It was not possible to use the extended theory of CI (Pishkin & Blanchard, 1963) to make error predictions or establish the value of social cues, since the equations apply only to the influence of programmed social cues. In conclusion, the mathematical theory of CI that bases its assumption on the performance of the individual working alone needs to be qualified or revised for pairs of learners free to exchange information.

REFERENCES

- ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *Journal of Experimental Psychology*, 1955, **49**, 155-164.
- BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *Journal of Experimental Psychology*, 1957, **54**, 201-207.
- BOURNE, L. E., JR., GOLDSTEIN, S., & LINK, W. E. Concept learning as a function of

- availability of previously presented information. *Journal of Experimental Psychology*, 1964, **67**, 439-448.
- BOURNE, L. E., JR., & HAYGOOD, R. C. The role of stimulus redundancy in concept identification. *Journal of Experimental Psychology*, 1959, **58**, 232-238.
- BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychological Review*, 1959, **66**, 278-296.
- CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology*, 1960, **59**, 137-144.
- ESTES, W. K. Learning theory and the new mental chemistry. In J. R. Braun (Ed.), *Contemporary research in learning*. New York: Van Nostrand, 1963. Pp. 202-229.
- HUNT, E. B. Memory effects in concept learning. *Journal of Experimental Psychology*, 1961, **62**, 598-604.
- LAUGHLIN, P. R. Selection strategies in concept attainment as a function of number of persons and stimulus display. *Journal of Experimental Psychology*, 1965, **70**, 323-327.
- PISHKIN, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *Journal of Experimental Psychology*, 1960, **59**, 371-378.
- PISHKIN, V., & BLANCHARD, R. J. Stimulus and social cues in concept identification of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1963, **67**, 454-463.
- PISHKIN, V., & WOLFGANG, A. Number and type of available instances in concept learning. *Journal of Experimental Psychology*, 1965, **69**, 5-8.
- RESTLE, F. A theory of discrimination learning. *Psychological Review*, 1955, **62**, 11-19.
- WOLFGANG, A., PISHKIN, V., & LUNDY, R. M. Anxiety and misinformation feedback in concept identification. *Perceptual and Motor Skills*, 1962, **14**, 135-143.
- ZAJONC, R. B. Social facilitation. *Science*, 1965, **149**, 269-274.

(Received March 19, 1966)

INTERACTIONS OF MODALITY WITH AGE AND WITH MEANINGFULNESS IN VERBAL LEARNING¹

J. C. COOPER, JR. AND J. H. GAETH

Wayne State University

58, 932 students of the Detroit Public School System, were used in a 3-factor experimental design to investigate the interactions among 5 grade levels, 2 modes of presentation (auditory vs. visual), and the learning of verbal materials at 2 levels of meaningfulness. A recall paired-associate paradigm was employed. Within the context of a single experimental program, heretofore unavailable, the presumed interaction of Modality $\times$ Grade Level and Meaningfulness has been confirmed. A significant interaction of Grade Level $\times$ Modality was found, although not in the expected direction. It is suggested that there is no inherently preferred modality for materials which are highly meaningful, that any such preference is a function of habit patterns.

A unique situation arises in the realm of verbal learning. A set of materials is available that can be readily converted from an auditory to a visual presentation and vice versa. With a great deal of generality, it may be said that an effect attributable to modality of presentation is reasonable only in a context in which alphabet materials are used. It is very difficult to create other sets of materials which could be called equivalent across modalities. Historically, questions concerned with the relative efficacy of auditory versus visual materials have produced answers which point to an interaction of modality with other factors. Day and Beach (1950, p. 8) reviewed modality considerations in verbal learning. As they point out, a limitation of the review was the inability to make direct comparisons among the studies available; task requirements, materials, and paradigms varied to a considerable extent. Certain conclusions were reached: (a) As the intelligence level and reading

level of the subject increases, the visual modality becomes more efficient. (b) "The relative efficiency of a visual presentation increases with age from a definite inferiority at the age of six to a possible superiority at the age of sixteen." (c) Increasing difficulty of material favors a visual presentation but "particularly easy material is better understood with an auditory presentation."

Recently, Williams and Derks (1963) concluded that modality of presentation produced a significant interaction with materials. The subjects were college students engaged in a paired-associate task. Keppel (1964) reviewed verbal learning in children with the aim of structuring experimental findings within current learning theory. Little information was uncovered with regard to the effects of modality of presentation. Gaeth (1960, 1963) has approached the problem of modality effects. The direct concern was the comparison and evaluation of differences in performance between normal and hearing-handicapped populations. In the first report, only modality of presentation and grade level were varied. No pattern of modality effects was evident, although modality

¹ The present study is based, in part, on a portion of Cooperative Research Project #1001, Office of Education, United States Department of Health, Education, and Welfare.

interacted significantly with grade level. In the second study, the ranges of grade level and type of material were expanded. A portion of that data provides the uniform testing conditions for a direct confirmation of effects of modality over a range of maturational levels.

METHOD

Design

A three-factor experimental design was employed. Five grade levels (4, 5, 6, 10, and 12) were represented. Two levels of meaningfulness (consonant-vowel-consonant trigrams and three-letter nouns) and two types of presentation (visual and auditory) were used.

Subjects

The subjects were 932 students from the fourth, fifth, sixth, tenth, and twelfth grades in the Detroit Public School System. These were randomly selected from within schools which in turn were selected, with the aid of the system's administration, as a typical cross-section of middle-class urban schools.

Apparatus

A slide-projector presented visual material, and a dual-channel tape recorder, auditory material. The second channel of a tape contained signals which activated the slide-changing mechanism of the projector.

Materials

Table 1 presents the specific items used as stimuli and responses at each of the two levels of meaningfulness. In addition,

TABLE 1
STIMULUS AND RESPONSE PAIRS FOR
EACH TYPE OF MATERIAL

CVC	Stimuli association value	Nouns	Responses
WUB	21	CAT	H
KEZ	29	ICE	M
JID	35	ARM	V
DAQ	38	RUG	L
WOJ	13	SKY	X
ZEG	25	GUN	F
$\bar{X}$	26.8		

Archer's association values for the CVC nonsense syllables are listed beside the specific nonsense syllables, as is the mean for the list. It will be noted that the set of response items is the same for the two levels of stimulus meaningfulness.

Procedure

The recall modification of a paired-associate paradigm was used. On a visual learning trial, the stimulus was presented for 2.5 seconds followed by the response for equal duration. The interpair interval was 5.5 seconds. The intertrial interval was approximately 15 seconds. A test trial followed the same procedure except for the omission of the response. The subjects were instructed to write the appropriate response, if they had learned it, during the 5.5-second interstimulus interval of a test trial. A learning trial was given first, followed by a test trial, another learning trial, and so on until 10 learning and 10 test trials had been completed. The order of each trial was randomized. In the auditory presentation a single pronunciation of a stimulus or a response was made in lieu of a visual projection. Timing remained the same. Testing was done in classroom sections.

RESULTS

The mean correct performance for 10 test trials of the 20 conditions can be seen in Figure 1. A three-way analysis of variance was computed using Winer's (1962) model for unequal cell frequency based on unweighted means. The three main effects were significant: grades, $F(4, 912) = 90.08, p < .01$; materials, $F(1, 912) = 213.52, p < .01$; modalities, $F(1, 912) = 10.94, p < .01$. Two interactions were significant, Modality $\times$ Material, $F(1, 912) = 4.12, p < .05$; Grade $\times$ Modality, $F(4, 912) = 3.35, p < .05$. The remaining two-way interaction, Grade $\times$ Material, and the three-way interaction were not significant. The significant main effects expected for grade level and type of material were found. Of particular interest is the significant main effect of modality as well as the significant interaction of Modality $\times$ Materials

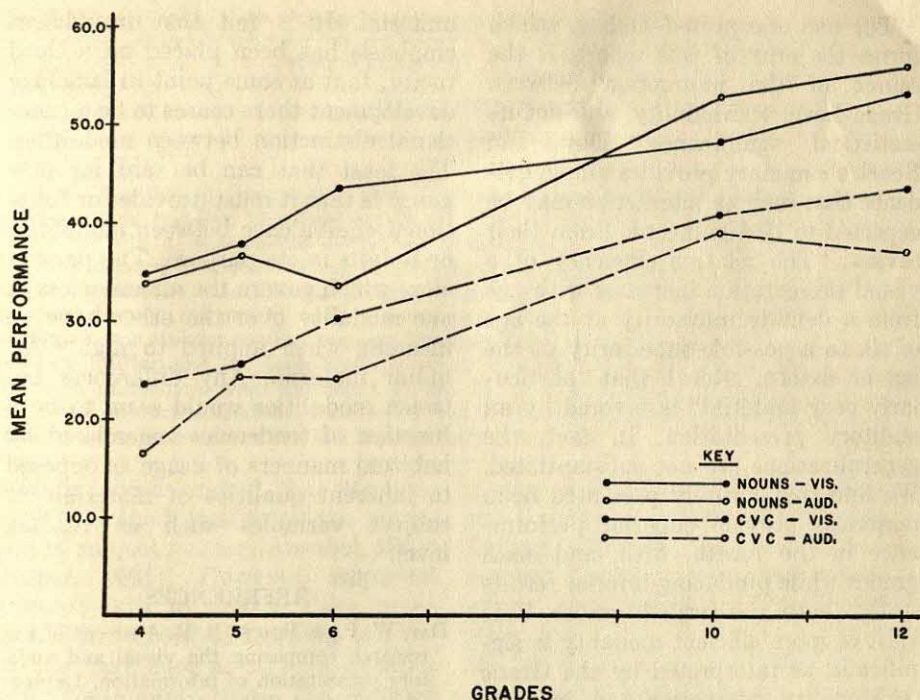


FIG. 1. Mean performance for 10 trials plotted against grade level for each modality of presentation and type of material.

and Grade Level, respectively. The significant interactions may be studied graphically in Figure 1. As can be seen, performance with nouns is superior with a visual presentation in the fourth grade, while an auditory presentation is superior in the twelfth grade. For trigrams, a visual presentation is consistently more efficient.

DISCUSSION

To be sure, the majority of the effects which can be inferred from the analysis of variance are completely in line with previous research. The main effects of grade level, meaningfulness of material, and modality fall into this category, although the effect of modality is not as clearly established in the literature as are the other two parameters. The three-way interaction is not significant, indicating that

we are not dealing with an overly complex situation. Nor would an interaction of Grade $\times$ Meaningfulness be expected. The interaction of Meaningfulness $\times$ Modality has not been previously examined in the context of a single experiment. However, its statistical significance, as found in this examination, would be expected from the review made by Day and Beach. In examining Figure 1, it is clear that the more difficult CVC material lends itself to visual presentation across grades, which is the converse of the statement that "particularly easy material" is favored by an auditory presentation. It will be noted that no hint of the Grade $\times$ Modality interaction can be seen with this type of material. The noun materials demonstrate no such uniformly more effective modality.

The one unexpected finding, which forms the crux of this report, is the source of the interaction between Grade Level $\times$ Modality, and not its statistical significance. Day and Beach's summary provides ample evidence that such an interaction may be expected to be significant. From their review, "The relative efficiency of a visual presentation increases with age from a definite inferiority at the age of six to a possible superiority at the age of sixteen." Recall that "particularly easy material" is favored by an auditory presentation. In fact, the generalizations are not substantiated. We find that visually presented noun materials provide superior performance in the fourth, fifth, and sixth grades while producing inferior results in the tenth and twelfth grades. This shift of more efficient modality is significant, as interpreted by the Grade $\times$ Modality interaction and, in addition, is in a direction opposite to that suggested by Day and Beach.

The basis for Day and Beach's prediction would seem to be their dependence upon two determining factors, the reading level of the subject and the nature of the materials. In the first case it is presumed that the greater degree of proficiency in and experience with oral communication would induce in the younger child a disposition toward oral communication which would lessen as reading ability and experience increased. In the second case the advantages of a visual presentation, its stability in time, its discreteness, and its inherent numerical advantage in terms of manipulatable parameters (versus the number available in an auditory presentation), lend themselves to communication of less familiar or more difficult

material. It is felt that insufficient emphasis has been placed on a third factor, that at some point in language development there ceases to be a functional distinction between modalities. The least that can be said for language is that it must provide for functional equivalence between modalities or it fails in its purpose. The parameters which govern the effectiveness of one modality over the other have no meaning when applied to highly familiar material. Any differences between modalities would seem to be a function of tendencies engendered by habitual manners of usage as opposed to inherent qualities of materials or subject variables such as reading level.

REFERENCES

- DAY, W. P., & BEACH, B. R. A survey of the research comparing the visual and auditory presentation of information. Dayton, Ohio: United States Air Force Base, Air Materiel Command, Wright-Patterson Air Force Base, 1950. (Air Force Technical Report 5921.)
- GAETH, J. H. Verbal learning in children with reduced hearing acuity. Final Report, Cooperative Research Project #289, Office of Education, United States Department of Health, Education, & Welfare, 1960.
- GAETH, J. H. Verbal and nonverbal learning in children including those with hearing losses. Final Report, Cooperative Research Project #1001, Office of Education, United States Department of Health, Education, & Welfare, 1963.
- KEPPEL, G. Verbal learning in children. *Psychological Bulletin*, 1964, **61**, 63-80.
- WILLIAMS, J. M., & DERKS, D. L. Mode of presentation and the acquisition of paired-associates that differ in pronounciability and association value. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 453-456.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received March 23, 1966)

EXPOSITORY INSTRUCTION VERSUS A DISCOVERY METHOD¹

JOHN T. GUTHRIE

University of Illinois

To test the proposition that discovery learning facilitates retention and transfer, 72 college seniors were taught to decipher cryptograms with 4 instructional sequences: Rule-Example, Example-Rule, and Example. A No-Training control group was included. On the retention test, the Rule-Example group was superior to all groups and the other groups did not differ. On the remote transfer test, the Example group was superior to all others, the Example-Rule and No Training did not differ, and the Rule-Example group was inferior to the others. All differences were significant at the .05 level. The discovery method appears to facilitate transfer, but not retention; expository instruction facilitates retention, but impedes remote transfer.

It has been argued that the discovery method facilitates the retention of subject matter (Ausubel, 1963; Bruner, 1961). However, empirical research fails to sustain these opinions. When speed of learning and retention were used as criteria, instruction containing rules has proved superior to instruction without rules (Craig, 1956; Haselrud & Meyers, 1958; Kittell, 1957; Wittrick, 1963). In addition, the only two studies (Hendrix, 1961; Katona, 1940) which oppose this generalization are difficult to interpret because the control and experimental groups were incomparable.

A more popular stand is that the discovery method is primarily useful for producing transfer and general problem-solving ability (Bruner, 1961; Suchman, 1961). However, when transfer is used as the criterion for learning, the results have been equivocal. Craig (1956) found no difference between "independent" and "directed"

groups on transfer to new rules. Haselrud and Meyers (1958) confirmed this finding, although this conclusion is suspect due to uncontrolled practice effects resulting from a within-subjects design. An advantage on transfer for instruction with "rule given" as compared with "rule not given" is reported by Wittrock (1963). However, the present author's examination of the rule group and the no-rule group with feedback present for both, indicates that there was an advantage ($t = 2.50$, $df = 145$, $p < .05$) for the no-rule group. Gagné and Brown (1961) also suggest that a "guided discovery" training procedure is superior to a "rule and example" procedure for producing transfer. Their treatment conditions, however, appear to have favored the "guided discovery" group. This group was trained on the specific type of behavior required by the criterion, whereas the "rule and example" group was taught a different type of behavior. Finally, Kittell (1957) concluded that giving a rule at the outset of instruction was superior to not giving a rule on all criteria: speed of learning, retention, and transfer. His conclusion, however, is unjustified since a substantial number of the subjects (Ss) taught by the discovery

¹This paper was submitted in partial fulfillment of the requirements for the degree of Master of Arts in Education at the University of Illinois. The author is indebted to Richard C. Anderson, who served as advisor for this thesis. The assistance of Elsa Roy in running subjects is acknowledged with appreciation.

method failed to meet the initial learning criteria. It is doubtful that Ss would be able to transfer behavior which was never learned.

The above research clearly demonstrates that instruction with rules produces faster learning and better retention than instruction without rules. On the other hand, the empirical evidence concerning techniques for producing transfer is inconclusive. Consequently, the more central proposition that discovery learning facilitates transfer remains virtually untested. The purpose of the present experiment was to compare several methods of instruction including both rules and examples and only examples on the basis of several criteria of retention and transfer. It was hypothesized that instruction with rules and examples would facilitate retention, but not transfer. And training with only examples would facilitate transfer, but not retention.

METHOD

The Ss were 72 students enrolled in an undergraduate educational psychology course. All Ss were taught to decipher cryptograms. The cryptograms were taken from words 4-10 letters in length occurring 20-30 times per million in adult reading matter (Thorndike & Lorge, 1944). The letters were scrambled according to one of six rules. Rules 1, 2, and 3 were transpositional; 4, 5, and 6 were substitutional. Rule 1: Exchange the first and the last letters in each cryptogram. Rule 2: Reverse the order of the first half and the last half of the letters. Rule 3: Remove the first 2 letters and place them in the middle of the rest of the cryptogram. Rule 4: Replace the first letter with the letter succeeding it in the alphabet. Rule 5: Replace the numbers with the correct vowel (a, e, i, o, u equal 1, 2, 3, 4, 5). Rule 6: Replace the last letter with the letter preceding it in the alphabet.

There were four treatment conditions to which Ss were assigned at random. In the Example-Rule group, examples of cryptograms were presented until a criterion of eight consecutive correct responses was attained. The rule was then taught with a

programming technique until the Ss could verbalize it upon request. Under the Rule-Example condition, the Ss were first taught the rule through the program. Examples of cryptograms governed by the rule were then given until the criterion was met. The Example condition consisted of presenting only examples of cryptograms until the criterion was attained. The control group received no training on deciphering cryptograms, but spent a comparable amount of time learning Russian vocabulary. All groups were subdivided. Half the Ss in each group were taught two transpositional rules (1 and 2); and half were taught two substitutional rules (4 and 5).

The test was composed of 30 cryptograms, and was subdivided into 3 parts. The first part contained 10 cryptograms formed from words not seen in training and governed by rules not used in training. The rules for this part were drawn from the opposite class of rules used in training. For instance, if S was trained to solve cryptograms composed from transpositional rules, this part required him to solve cryptograms composed from substitutional rules, and vice versa. This was considered the "remote" transfer task. The second part contained 10 cryptograms formed from words not seen in training, but governed by exactly the same rules as used in training. This was considered the retention task. In the third part, the 10 cryptograms were formed from new words and governed by rules not used in training. However, these rules were drawn from the same class of rules (transpositional vs. substitutional) as used in training. This was labeled the "near" transfer task.

The Ss were seated at a table facing an upright divider ($2\frac{1}{2} \times 5$ feet) with a small centered window (3×5 inches) in which the stimuli were presented. Written instructions were first given to each S on a sheet of paper. The instructions, which were minimal, directed the S to say the word suggested by each cryptogram and informed him that he would be tested. The instructions for the control group pertained only to the Russian vocabulary.

The Ss received successive presentation of cryptograms conforming to one of the rules. Each S was presented cryptograms until a criterion of 8 consecutive correct responses was attained or until 25 cryptograms had been presented. The order of presentation was randomized for each S. Each trial consisted of presenting a cryptogram and requiring S to verbalize the correct word. The length of the trials was 15 sec-

onds. Feedback, which consisted of presenting the cryptogram with the correct word beside it, followed each trial. The duration of feedback was 5 seconds. The time lapse between feedback and the following trial was about 5 seconds. Each *S* was taught to decipher cryptograms governed by two rules. After the two rules were learned, the test was given. Directions for the test informed the *S* of the test and stated that some of the correct answers were obtainable by substituting one letter for another letter in the alphabet, and other answers were obtainable by rearranging the letters within the cryptogram. The *S* was informed when he entered each of the test subdivisions. The approximate time required for the test was 10 minutes.

RESULTS

The data were initially analyzed with a series of orthogonal planned comparisons, but the planned comparisons appeared too general to adequately represent the data. Consequently, simple one-way analyses of variance were computed for each of the test subdivisions including retention, near transfer, and remote transfer and for the trials to criterion. The means, standard deviations, and *F* ratios are presented in Table 1.

The overall significant differences permitted the use of the Tukey procedure (Edwards, 1963) for comparing

means (see Table 2). First, on the remote-transfer task, the Example group was significantly superior to all other groups. The Example-Rule group did not differ from No Training. And the Rule-Example group was significantly inferior to all other groups. On the near-transfer task, the Example group and the Example-Rule group were superior to No Training and Rule-Example, but did not differ from each other. The Rule-Example group was not different from the controls.

On the retention task, there do not appear to be any differences among the treatment groups, though all are superior to no training. However, for reasons outlined in the discussion, it is believed that the retention test for those trained on substitutional rules was incapable of discriminating among treatments. Consequently, analyses of variance were conducted separately for *Ss* trained on substitutional rules and *Ss* trained on transformational rules. Significant differences were found on both groups (see Table 1). The Tukey procedure was then employed to compare the means (see Table 2). As expected, there were

TABLE 1
MEANS AND STANDARD DEVIATIONS OF ERRORS AND TRIALS TO CRITERION

	Rule-example		Example-rule		Example		No training		<i>F</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Remote transfer	5.94	1.59	4.26	2.29	2.56	2.73	4.67	1.15	6.64**
Near transfer	5.23	2.26	2.84	2.00	2.27	1.05	4.50	1.21	8.94**
Retention	1.94	2.37	3.06	2.31	2.89	2.42	5.27	2.27	7.64**
Retention (Substitutional)	1.33	1.70	1.56	1.06	1.11	0.29	4.44	2.64	7.13**
Retention (Transformational)	2.55	1.37	4.56	2.26	4.67	2.25	6.11	1.41	4.80**
Trials to criterion	11.92	6.96	23.66	8.51	22.72	8.81	—	—	4.40*

Note.—All test data are based on errors. Trials to criterion represent the total trials required to learn two rules.

* $p < .025$.

** $p < .01$.

TABLE 2
TUKEY COMPARISON OF MEANS

Remote transfer	Near transfer	Retention	Retention substitution	Retention transformation	Trials to criterion
1. Example	Example	Rule-Example	Example	Rule-Example	Rule-Example
2. Example-Rule	Example-Rule	Example	Rule-Example	No training	Example
3. No training	No training	Example-Rule	Example-Rule	Example-Rule	Example-Rule
4. Rule-Example	Rule-Example	No training	No training	Example	—

Note.—Treatments are in rank order. The treatments connected by a common line do not differ. All other treatments differ significantly at the .05 level.

no differences among groups trained on the substitutional rule, though the control group was inferior to all treatment groups. However, for groups trained with transformational rules the Rule-Example was superior to all other groups. In addition, the Example and Example-Rule groups did not differ from the No-Training group.

To analyze the time required for learning under various conditions, a one-way analysis of variance was conducted on all treatment groups using trials to criterion as the dependent variable. First, the general differences were significant. In addition, the Tukey comparison showed that the Rule-Example group learned faster than the other two groups; and that the Example and Example-Rule groups did not differ from each other.

DISCUSSION

At the outset of the experiment, it was hypothesized that the rules would function as eliciting stimuli. In other words, it is suggested that rules in verbal learning are analogous to the unconditioned stimuli in classical conditioning. They serve to evoke a response which is then reinforced. For example, when the rule was presented in the Rule-Example group and a cryptogram was presented in close succession, specific deciphering behavior was evoked from Ss. When these Ss were presented a cryptogram,

they manipulated the letters specifically according to the rule to form the word. It was speculated that this behavior would then be reinforced through feedback and acquired. That such behavior was, in fact, acquired is evident from the results indicating that Ss in the Rule-Example group performed well on retention tasks.

It was also reasoned at the outset that when no rules were given to Ss, their behavior would more closely approximate the operant conditioning paradigm. That is, presenting Ss a cryptogram would evoke searching, exploratory behavior which would be reinforced when a word was made from the cryptogram. That such explorative skill was acquired is evidenced from the result that the Example and Example-Rule groups surpassed the Rule-Example and the Control groups on the transfer, but not on the retention tasks.

It should be noted that the conclusions regarding retention have been drawn from the data of Ss taught transformational rules; and the data from Ss taught substitutional rules have been ignored. The reason for this is that one of the substitutional rules contained a number, whereas all other rules contained only letters. It is likely that this number acted as a distinctive cue which facilitated retention in all groups. Thus a ceiling effect for retention of Ss taught substitutional rules prevented differences

on this measure from manifesting themselves.

REFERENCES

- AUSUBEL, D. P. *The psychology of meaningful verbal learning*. New York: Grune & Stratton, 1963.
- BRUNER, J. S. The act of discovery. *Harvard Educational Review*, 1961, **31**, 21-32.
- CRAIG, R. C. Directed versus independent discovery of established relations. *Journal of Educational Psychology*, 1956, **47**, 223-234.
- EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Holt, Rinehart & Winston, 1963.
- GAGNÉ, R. M., & BROWN, L. T. Some factors in the programming of conceptual material. *Journal of Experimental Psychology*, 1961, **62**, 313-321.
- HASLERUD, G. N., & MEYERS, S. The transfer value of given and individually derived principles. *Journal of Educational Psychology*, 1958, **49**, 293-298.
- HENDRIX, G. Learning by discovery. *Math Teacher*, 1961, **54**, 290-299.
- KATONA, G. *Organizing and memorizing*. New York: Columbia University Press, 1940.
- KITTELL, J. E. An experimental study of the effect of external direction during learning on transfer and retention of principles. *Journal of Educational Psychology*, 1957, **48**, 391-405.
- SUCHMAN, J. R. Inquiry training: Building skills for autonomous discovery. *Merrill Palmer Quarterly of Behavioral Development*, 1961, **7**, 148-149.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.
- WITTRICK, M. C. Verbal stimuli in concept formation: Learning by discovery. *Journal of Educational Psychology*, 1963, **54**, 183-190.

(Received April 21, 1966)

EFFECT OF CURIOSITY ON INCIDENTAL LEARNING¹

WILLIAM PARADOWSKI

Teachers College, Columbia University²

Curiosity was aroused in an S by the presentation of 5 illustrations of strange-looking animals. By contrast, to determine the effect of low arousal, 5 illustrations of familiar animals were shown to the same S. Paragraphs of verbal information were paired with each illustration and provided an intentional learning task. 52 undergraduate Ss were used. Intentional learning was assessed for each item as was incidental learning, which consisted of the posttest recall of the settings and border colors around each animal illustration. Curiosity arousal—stimulated by the pictures of novel animals—significantly increased both intentional learning and incidental learning. This finding is at sharp variance with the generalization based on research on aversive drive states that drive arousal leads to reduced incidental learning.

A recently proposed generalization maintains that drive arousal acts consistently to reduce the range of cue utilization. According to Easterbrook (1959), the range of cue utilization is said to have fallen when the amount of incidental learning has been reduced although task learning has remained constant or been improved. A growing number of experiments have demonstrated this principle using a variety of drive states—for example, hunger (Bruner, Matter, & Papanek, 1955), thirst (Johnson, 1952), and anxiety (Silverman, 1954)—of which all can be classified as aversive in nature. In proposing his generalization, Easterbrook considered drive as of aversive character, and defined it as an innate response to a state of biologic deprivation or noxious stimulation.

Research on the tendency for aroused aversive drives to reduce incidental learning has been moderately consistent in its positive findings. No

investigation, however, has so far been undertaken to see if a fundamentally nonaversive drive would show the same tendency—as, for example, curiosity, a drive characterized by positive affect and approach behavior. There seems little reason to expect that with regard to range of cue use an organism aroused by curiosity would necessarily function in the same way as when aroused by an aversive drive.

For at least the past decade a number of theorists have emphasized the importance of recognizing a fundamentally different class of drives or motives than those classed together as aversive or as dependent on the homeostatic regulation of organic need. Woodworth (1958), for example, proposed that there are two independent classes of motives: need-primacy (which refers to motives primarily directed to the satisfaction of the internal needs of the organism) and behavior-primacy (which refers to motives directly concerned with the interaction between the organism and environment). White (1959) has supported a similar view. Hebb (1955) also has made a restatement concerning different effects of drive on learning and motivation. Referring specifically to curiosity, some educators have conceived of it as a condition

¹ Adapted from a dissertation submitted to Teachers College, Columbia University in partial fulfillment of the requirements of the doctoral degree in clinical psychology. Laurance F. Shaffer, Rosalea A. Schonbar, and Walter H. MacGinitie served as members of the dissertation committee.

² Now with Psychological Services, Goldwater Memorial Hospital, Welfare Island, New York.

whereby an individual becomes more sensitive to cues in his environment (Maw & Maw, 1962).

Therefore, it is hypothesized, first, that arousal by curiosity has an effect on incidental learning different from that evoked by low arousal. Second, evidence is sought that would support a proposition that (a) curiosity arousal increases incidental learning or, alternatively, that (b) curiosity arousal decreases incidental learning.

METHOD

The experimental design followed the model of incidental learning studies in which intentional and incidental tasks are so incorporated into one learning situation that the effects of a high- and low-arousal condition may be compared for each task. A booklet of written and illustrated material was prepared to conform to this design. Pictures of animals were paired with paragraphs of information about the animals so that the written matter provided an intentional task and details in the illustrated matter provided an incidental task.

Since odd or strange-looking animals are a frequent source of curiosity, it was planned that pictures of such animals would be used—in contrast to pictures of quite familiar animals—to achieve conditions of high and minimal curiosity. The pictures of the animals used in the experiment proper were chosen on the basis of an auxiliary study. Twenty-four animal illustrations were prepared by the experimenter, 12 of which had been selected for their novelty and 12 for their ordinary appearance. Twenty subjects, (Ss), employed in the auxiliary group, were asked to rank the pictures of the animals from most interesting to least interesting. From the results, five animals consistently ranked as most interesting were selected as the experiment's "novel" animals, and five animals consistently ranked as least interesting were chosen to be the "common" animals. Of the novel animals, the one with the highest rank position for interest was an imaginary creation with a mean rank position of 4.55. The other four novel animals represented actual species: a pichichiago (5.25); a megatherium (5.45); an elephant shrew (6.35); and a hemigale (6.80). The common animals, with their mean rank positions, included: a donkey (18.60); a deer (18.95); a

dog (19.80); a sheep (20.00); and a muskrat (20.60).

Since the stimulus for curiosity was to be an animal illustration, it seemed practical to provide a background to the animal that could serve as a source of incidental learning material. With this in mind, 10 easily identifiable environmental settings were drawn. These comprised five types of environment with each type represented by two different drawings. The five types included a desert, forest, jungle, swamp, and one of mountainous terrain. Recall of a setting represented an item of incidental learning. In addition, 1-inch colored borders were placed around each picture, five colors being used, each repeated twice. Color recall provided a second kind of incidental learning item.

A valid comparison of the incidental learning achieved for the novel animals as opposed to the common animals hinged on the comparability of the incidental cues paired with each set of animals. Therefore, in order to equate the incidental learning task for the two conditions, two sets of pictures were prepared. In one set the five novel and five common animals were placed in particular environmental settings; in the second set the relation between animal and background was reversed. For example, if a novel animal appeared in a forest in Set A, then a common animal appeared in the same forest in Set B. This reversal procedure applied to colored borders as well as to environmental settings. Two groups of comparable Ss were used, one for each set of pictures. In this way, incidental learning could be assessed for all the settings and all colored borders—once when paired with a common animal and once when paired with a novel animal. Particular effort was made to avoid placing animals in environmental settings that might produce an incongruous effect.

The complete experimental instrument consisted of the described pictures presented together with written "factual" information about each animal. The material was presented in a large book, and arranged so that an illustration with a colored border appeared on the left-hand page and written matter on the right-hand page. The instrument was prepared in this way to resemble material that could conceivably have come from a textbook on vertebrate paleontology.

Two such books of material were prepared. The second book was identical in all respects to the first except that wherever a novel animal appeared in the one set of

illustrations, a common animal appeared in the other.

The *Ss* were told that they were to assist in the evaluation of a forthcoming general text for undergraduates by taking short tests on each of the presented items to assess the ease with which the material could be learned. This rationale provided the basis for the central task of the experiment. After looking at a picture of an animal and then reading a paragraph of "factual" material supposedly about the pictured animal, each *S* was given a completion test on the paragraph read. As far as *Ss* were concerned, their main purpose in studying the presented material was to do the completion tests. The content of the paragraphs was concerned with information of such a trivial nature that—to a naïve reader—it could have easily applied to any animal. None of the presented statements of information conflicted with known facts about familiar animals. Since two sets of pictures had been prepared, each paragraph was also used twice, once paired with a novel animal and once paired with a common animal. As an example, one of the paragraphs used in the study follows:

Fossil remains of this animal indicate an age of 36 million years. It has been determined to be of Miocene origin and is generally considered as a marginal species between the Tertiary and Quaternary periods. The only fossils of this species have come from central Africa. The feet of this species have remained essentially unchanged from its primitive predecessors as well as its phalangeal formula. The teeth too are essentially unchanged with a dental formula of five upper and four lower incisors, a canine, three premolars, and four molars. This species shows a relatively fast postnatal development whereby the young are independent of their mothers and capable of self-care at an early age.

A second auxiliary study was done to prepare a completion test on the "factual" material. It was desired that the completion test give a reliable measure of what the *Ss* had learned from the central task material in order to determine whether or not curiosity arousal could also affect the retention of that material for which a direct learning set was given. The paragraphs of information were reproduced with 12 word deletions and were presented to 20 *Ss* with the complete paragraphs in alternating order. The *Ss* were permitted to read each paragraph once and then were given the appropriate

completion test. Based on the results of this pilot study, the final completion test was composed of six deletions. Completion items were chosen that were answered correctly by approximately 50% of the auxiliary group. Thus, items that were too easy or too difficult were eliminated.

Procedure

Two groups of 26 undergraduates were used as *Ss*. They were oriented to the task by being asked to assist in evaluating material from a forthcoming textbook. The book of material was then opened by the experimenter who also controlled the presentation of each test item. A practice item was first shown to *Ss* which was identical in format to the succeeding 10 experimental items. While the written paragraph was covered, the illustrated material was exposed for 10 seconds. At the end of this time the written text was exposed for 20 seconds while the illustration was covered. The page was then turned, thus terminating exposure to the test material. A completion test for the paragraph read was immediately presented to the *Ss*. Following the completion items, a multiple-choice test was administered which was concerned solely with the appearance of the animal. (Such a test was necessary as a rationale for having the *Ss* look at the pictures. The scores, however, were not tabulated or in any way used in the final results.) Upon completing the two tests, *Ss* were presented with the next test item in its order. After all the items were done, *Ss* were given an incidental learning test which consisted of 10 pages, each page containing a reduced illustration of one of the experimental animals. Beneath each illustration were two multiple-choice questions. The first question called for the selection of the actual background of the depicted animal; the second required selection of the color used for the border of that animal's picture.

RESULTS

In addition to the auxiliary study ranking the animals for interest, a separate auxiliary study was done to demonstrate the preference of *Ss* for looking at pictures of the novel animals as opposed to common animals. Only 10 *Ss* were used but the results were clear. The procedure involved presenting the complete test material, as described, to *Ss*. Both the illustrated and written material, however, were

simultaneously exposed. The *Ss* were therefore free to either glance at the illustration or read the accompanying paragraph, as they wished. The amount of time *Ss* spent looking at the pictures was surreptitiously observed, timed by a stopwatch, and recorded. Within a 50-second time limit per item, *Ss* spent significantly more time looking at the pictures of novel animals (45.5 seconds) than at the pictures of common animals (31.1 seconds). A two-tailed *t* test showed a .04 probability value for the difference in time.

The findings of the auxiliary study can be applied to the experiment proper, although there were some differences in procedure. Instead of being free to divide their time for 50 seconds between the pictures and the paragraphs as in the auxiliary study, *Ss* were exposed to the pictures for 10 consecutive seconds. The *Ss* invariably spent this entire span of time examining the illustration. The exact amount of time an *S* looked at the animal in an illustration as opposed to its background cannot be determined. The findings of the auxiliary study, however, suggest that novel animals within their settings were attended to proportionally longer than common animals within their settings. This assumption had been further safeguarded by giving *Ss* a set for examining the animals within the illustration. The short series of multiple-choice questions (mentioned in the Procedure section) were concerned only with the appearance of the animal and in no way involved in the settings. A common animal, by virtue of its familiarity, is unlikely to sustain interest and attention. (For example, it only takes a moment to recognize and label a dog.) Therefore, it is assumed that *Ss* spent less time examining the backgrounds of a novel animal than of a common animal.

Intentional and Incidental Learning

Scores were tabulated according to their relation with novel or common animals. Thus each *S* had a novel condition learning score and a common condition learning score, both for incidental and intentional tasks. Novel scores—for all *Ss* in both groups—were compared to common scores. Since all extraneous factors between the two conditions had been controlled, direct statistical comparison was made by means of two-tailed *t* tests. The formula used for the *t* tests, however, took account of the size of the coefficient of correlation between the two sets of scores. The results are summarized in Table 1.

Although the intentional task scores covered a wide range under both conditions, they were nonetheless highly correlated. The substantial size of the correlation between the sets of scores (.77) is not surprising considering that all the completion tests were of comparable difficulty and that the same intellectual ability is invested in the tests under both conditions. It is not an unexpected finding, therefore, that *Ss* who scored high on the completion tests associated with novel animals also scored high on the completion tests associated with common animals.

A more interesting and unexpected finding, however, concerned the effect on intentional learning of curiosity arousal. A two-tailed *t* test shows a significant improvement in intentional task learning under the novel animal condition as opposed to the common animal condition.

The fact that the intentional task learning was significantly improved is a good indication that a state of arousal was operative. Improved central task performance is a standard criterion for increased motivation (Easterbrook, 1959). The *Ss* were better motivated to learn, although

the single fact of improved intentional task performance alone does not indicate the identity of the motivation. The nature of the motivation is perfectly clear, however, when the facilitated intentional learning is considered together with the results of the auxiliary study ranking the animals for interest and the auxiliary study demonstrating Ss' preference for looking at novel animals. All these findings demonstrate that Ss were motivated by curiosity, and constitute the operational criteria in this experiment for curiosity arousal.

The results on incidental learning clearly point to significant differences. The direction taken by the differences indicates that there is increased incidental learning with curiosity arousal.

TABLE 1

COMPARISONS OF INCIDENTAL AND INTENTIONAL RECALL OF MATERIALS PRESENTED WITH NOVEL AND COMMON ANIMALS

Variable	Means and SD		<i>t</i>	<i>r</i> ^a
	Novel	Common		
Intentional task recall ^{b**}				
<i>M</i>	14.75	13.58	2.72	.77
<i>SD</i>	3.85	4.77		
Total incidental recall ^{c**}				
<i>M</i>	4.21	3.25	4.05	.38
<i>SD</i>	1.67	1.37		
Setting recall ^{d**}				
<i>M</i>	2.81	2.37	2.75	.60
<i>SD</i>	1.27	1.33		
Color recall ^{d*}				
<i>M</i>	1.40	.88	2.36	-.07
<i>SD</i>	1.01	.73		

Note.—*N* = 52.

^a Correlation between recall scores for the "novel" vs. the "common" component of each variable.

^b A score of 30.0 is perfect for each condition.

^c A score of 10.0 is perfect for each condition. Chance expectancy for each condition is 2.0.

^d A score of 5.0 is perfect for each condition. Chance expectancy is 1.0.

* *p* < .05.

** *p* < .01.

The direction is maintained by both components of the total incidental learning score.

Curiosity arousal facilitated the recall of settings. The correlation coefficient for the recall of settings when paired with a novel animal versus recall when paired with a common animal was high, indicating that Ss who performed well under one condition tended to perform well under the other condition also. Thus, the coefficient of .60 may be considered an alternate-form reliability coefficient for five items of the recall test. By the Spearman-Brown formula, the probable reliability of 10 items would be .75.

Differences in recall were maintained by the color variable. Recall under the novel animal condition was significantly higher than the common animal condition, despite the fact that the two sets of scores showed virtually no correlation. Had the correlation been as high as the one existing between the novel and common conditions for the recall of settings, then a comparably high value for *t* would have been shown.

Since both components of the total incidental learning score showed significant differences in the same direction, it is to be expected that the total of these sets of scores would produce an even smaller probability value. Indeed, the *t* score for the differences between the novel and common conditions produced a probability value far beyond the .01 level. The novel animals thus appear to greatly facilitate total incidental learning.

The results of the study offer clear evidence of differential effect on incidental learning by curiosity arousal. Furthermore, curiosity arousal serves to increase incidental learning whether the cues are relevant or irrelevant to the eliciting stimulus.

DISCUSSION

The findings on the facilitation of intentional task learning were unexpected—indeed, somewhat surprising—since the intentional task material was not spacially congruent with the curiosity-eliciting stimulus. The picture of the novel animal was on the left-hand page while the paragraph of text was on the right-hand page. Moreover, the picture and the text material were never seen simultaneously. Despite the physical distance involved, however, Ss seemed to be motivated to examine the written material about the strange animals with more enthusiasm than that inspired by the familiar animals.

With regard to learning theory, some who espouse the Hullian approach (Myers & Miller, 1954) have felt that exploratory behavior, merely because it is externally elicited, should, like pain, be no differently treated than other drives. The present findings, however, indicate differences between curiosity and aversive drives, such as pain, with regard to incidental learning that call for different predictive formulations.

The effect on incidental learning of a variety of motivating conditions has been described by Easterbrook (1959) and again by Kausler and Trapp (1960). The results of the present study, however, suggest that the generalization proposed by Easterbrook concerning the relation between incidental learning and drive arousal may be meaningfully extended.

The hypothesis presents itself that the effects of arousal on incidental learning are dependent upon the nature of the drive aroused. Although

aversive or noxious stimulation tends to reduce the range of cue utilization, arousal conditions characterized by positive affective tone and interest may increase the range of cue utilization. Such a hypothetical generalization may be stated in Woodworth's (1958) terms as follows: Arousal of need-primacy motives decreases incidental learning, whereas arousal of behavior-primacy motives facilitates incidental learning.

REFERENCES

- BRUNER, J. S., MATTER, J., & PAPANEK, M. L. Breadth of learning as a function of drive level and mechanization. *Psychological Review*, 1955, **62**, 1-10.
- EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 1959, **66**, 183-201.
- HEBB, D. O. Drives and the C. N. S. (conceptual nervous system). *Psychological Review*, 1955, **62**, 243-254.
- JOHNSON, E. E. The role of motivation strength in latent learning. *Journal of Comparative and Physiological Psychology*, 1952, **45**, 526-530.
- KAUSLER, D. H., & TRAPP, E. P. Motivation and cue utilization in intentional and incidental learning. *Psychological Review*, 1960, **67**, 373-379.
- MAW, W. H., & MAW, E. W. Children's curiosity as an aspect of reading comprehension. *Reading Teacher*, 1962, **15**, 236-240.
- MYERS, A. K., & MILLER, N. E. Failure to find a learned drive based on hunger; evidence for learning motivated by "exploration." *Journal of Comparative and Physiological Psychology*, 1954, **47**, 428-436.
- SILVERMAN, R. F. Anxiety and the mode of response. *Journal of Abnormal and Social Psychology*, 1954, **49**, 538-542.
- WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, **66**, 297-333.
- WOODWORTH, R. S. *Dynamics of behavior*. New York: Holt, 1958.

(Received April 1, 1966)

SELECTIVE FACILITATIVE EFFECTS OF INTERSPERSED QUESTIONS ON LEARNING FROM WRITTEN MATERIALS

ERNST Z. ROTHKOPF AND ETHEL E. BISBICOS¹

*Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey*

Does the use of restricted categories of questions, incorporated in written instructional material, facilitate learning of restricted categories of text content? High school students ($N = 252$) saw a 36-page passage with 2 questions per 3-page zone. Questions differed in location (before or after the relevant segment) and in required response. Different treatments saw questions restricted to the following answer types: (a) either a quantitative term or name; (b) a common English or a technical word; (c) a mixture of (a) and (b). The experimental hypothesis was confirmed. Learning of several categories of text content, as measured on a posttraining retention test, was facilitated by appropriate questions seen immediately after exposure to the relevant text segment. The results were interpreted to be a consequence of modification of inspection behavior.

Several experimental investigations have shown that, under certain conditions, the incorporation of questions and other test-like features in written, instructional material increased the amount learned from the text (e.g., Hershberger, 1964b; Keislar 1960). Rothkopf (1963, 1965, 1966), and Rothkopf and Coke (1963) have reported that the facilitating effects of questions have at least two components. These are: (a) direct instructive effects, that is, questions are informative; and (b) general, attention-like effects, that is, effects on the inspection (studying) behavior of the student. The second effect was observed to be a function of where questioning occurred in the stimulus materials. Periodic questions about material which had already been seen were found to result in more effective inspection (or mathemagenic)² be-

haviors than periodic questions about matter which subjects (Ss) were about to study.

The purpose of the present experiment was to find answers to the following: (a) Does the use of restricted categories of questions in instructive text result in inspection behaviors which facilitate learning of restricted categories of text content? (b) Do facilitative effects which are attributable to inspection behaviors change as a function of increased exposure to the experimental questions?

A technique which had been developed in an earlier study (Rothkopf, 1966) was used to separate direct informative effects of questions from effects on inspection behavior. A number of short-answer questions were composed for a lengthy factual passage. These questions were divided into two equal subsets, A and B. Subset A was used as experimental questions during study. Subset B was administered after training to measure how much was learned. However, the two subsets of questions were selected so that there was no observable direct transfer of training from mastery of the subject matter on which Subset

¹ We wish to thank our colleague Esther Coke for her help in the computer analysis of the data.

² The roots of this useful, but long word are *mathema*: learning, that which is learned and *gignesthai*: to be born. This seems an appropriate reference to a class of responses which give birth to learning (see Rothkopf, 1965).

A was based to the subject matter from which Subset B was derived. It was assumed that experimental verification of this relationship between A and B implied that any changes in Subset B performance, which result from exposure to the questions of Subset A, were not due to the informative aspects of questions of Subset A.

METHOD

Materials

Two chapters from Rachel Carson's book, *The Sea Around Us*, were used.^{*} The experimental passage was approximately 9000 words long. It dealt with animals and minerals found in the ocean. The material, although topically related, was composed of a series of relatively independent factual segments.

The experimental passage consisted of 36 typewritten pages which were divided into 12 three-page zones. Eight short-answer questions were composed for the material in each three-page zone. The eight questions for each zone could be subdivided into four categories of two questions each according to the character of the correct answer. The categories were: (a) common phrases (C), any common nontechnical English words, for example, red or sense of touch; (b) technical phrases (T), scientific or technical words, for example, bathyscaphe or phototrophic, all of these were of either Greek or Latin origins and could be broken into meaningful component roots; (c) measures (M), quantities of size, numerosity, distance, as well as dates, for example, 1/2 ton or 1942; and (d) names, (N), any proper geographical or personal names or any other name-like word which had been assigned more or less by fiat, for example, Johnson, Philippines, or Jesuit. A sample question for each of the above categories is given below.⁴

^{*}Permission for the experimental use of these copyrighted materials was kindly granted by the publishers, Oxford University Press, 417 Fifth Avenue, New York 16, New York.

⁴The correct answer, and its classification according to experimental type, follows each sample question in parenthesis. (a) The small whale which is popularly called a porpoise in America is really the _____

Four questions from each zone, one from every answer category, were used to make up the 48-item criterion test. This test was administered after the completion of training and was used for the main experimental comparisons. The remaining 48 questions (i.e., four questions per zone) were used as the experimental questions (EQs) which were located in various portions of the stimulus material.

Experimental Treatments

The seven experimental treatments are summarized in Table 1. The NOEQ group saw no experimental questions. All of the remaining treatments saw two experimental questions per three-page zone, but differed in type and location of experimental questions. Two of the treatments (SBNM and SANM) saw identical sets of experimental questions all of which required measured quantities (M) or names (N) for answers. For SBNM, however, the two experimental questions about a given zone were presented immediately before the three-page zone was to be read (SB = shortly before). While for SANM, experimental questions were presented immediately after (SA = shortly after) the zone from which they were taken. In a like manner, SBCT and SACT were exposed to the identical set of experimental questions, all of which required either a common (C) or a technical word (T) or phrase for a correct answer. But in SBCT the experimental questions preceded their zone, while for SACT they followed. Finally SAMX (shortly after-mixed) and SBMX (shortly before-mixed) received the same set of experimental questions which was composed of a mixture of questions of all four types, selected so that each of the four types of experimental questions was used exactly three times in each half of the experimental text (six zones). The experimental orders for SAMX and SBMX were arranged so that each of the 48 questions in

dolphin (bottle-nosed; common phrase). (b) The decaying organic matter on the bottom of the stagnant Norwegian fiords produces the chemical called _____ which poisons the deeper waters (hydrogen sulphide; technical phrase). (c) The concentration of bromine in the Dead Sea is _____ times as great as it is in the ocean (100; measure). (d) The building of the bathyscaphe was first proposed by the Swiss physicist, Professor Auguste _____ (Piccard; proper name).

TABLE 1
EXPERIMENTAL DESIGN

Treatment	Number of experimental questions per three-page zone	Type of experimental questions used	Location of experimental questions relative to relevant three-page zone
SBCT	2	C, T	shortly before
SBNM	2	N, M	shortly before
SBMX	2	C, T, N, M	shortly before
SACT	2	C, T	shortly after
SANM	2	N, M	shortly after
SAMX	2	C, T, N, M	shortly after
NOEQ	0	—	—

the experimental question pool was used an equal number of times over all Ss.

Procedure

The experimental procedure was identical to that used in an earlier study and has been described in detail elsewhere (Rothkopf, 1966). All experimental materials, including the directions for their use, were bound in a loose-leaf notebook which was handed to *S* as he entered the room in which the experiment was conducted. The Ss entered individually and worked at one of the 10 tables which were widely spaced in a 21- × 60-foot room. Each table was partially screened from the view of adjacent Ss.

Every *S* worked through the materials at his own rate. They were instructed to record the time they started and finished each page from a direct-reading clock located at their work table. When *S* finished reading a page of text or an experimental question, he tore it from the notebook and dropped it through a slit into a ballot-box-like container.

The criterion test was administered immediately after training. Four forms of the test were used. These differed only in the sequence of the 48 questions.

Subjects

Paid volunteers ($N = 252$) from Grades 10, 11, and 12 of Governor Livingston Regional High School of Berkeley Heights, New Jersey,⁵ served as Ss immediately

⁵ We wish to express our sincere thanks to W. M. Davies, Superintendent, Union County Regional High School District No. 1, as well as to F. Aho, Principal, and J. S. Stein, both of Governor Livingston High School, for their help in procuring student volunteers for this study.

after the end of their regular school day. Thirty-six were assigned in a nonsystematic manner to each of the seven experimental treatments. A group of 23 additional Ss were used to evaluate the direct instructive (transfer) effects of EQs on the skills required for high criterion test performance.

RESULTS

The absence of transfer between subject matter skills inducing high performance on experimental questions and those underlying performance on criterion test questions were evaluated by the following procedure. A group of 23 Ss, from the same population as the experimental groups, practiced giving the correct answers to experimental questions by a modification of the anticipation method (for details see Rothkopf, 1966). Following mastery of all questions, one of the forms of the criterion test was administered. The scores were compared with those of another form of the criterion test which had been administered prior to practice on the set of experimental questions.

The mean of the observed pretraining scores was 3.17 correct responses. The posttraining average was 3.04. The net change of $-.13$ offers no ground for rejecting the hypothesis that the skills which are required to answer the experimental questions *do not* transfer to performance on the criterion test. These findings indicate that any improvement in criterion test performance associated with exposure to experimental questions was not due to the direct informative effects of these questions.

Questions on the criterion test were divided into four numerically equal categories according to a 2×2 table with the following classifications: Factor 1, correct answer type; (a) answer either measures (*M*) or proper names (*N*), or (b) answer either common words (*C*) or technical phrases (*T*). Factor 2, location in the experi-

mental text of the material on which the question was based, that is, either: (a) pp. 1-18, or (b) pp. 19-36. Mean correct responses to criterion test questions of the two classes of answer types are plotted as function of location of text source, for the seven experimental treatments, in Figure 1. These results may be summarized as follows:

Treatments in which experimental questions were administered after exposure to the relevant text segment resulted in better overall retention test performances than the corresponding treatments in which experimental questions were administered just before inspection of the relevant test segment ($SANM > SBNM$, $SACT > SBCT$, $SAMX > SBMX$). This effect is weak for $SACT$. The administration of experimental questions after exposure to the relevant text segment produced higher retention test performance than the $NOEQ$ condition. But experimental questions just prior to inspection of the appropriate text segment resulted in performance which was not distinguishable from this control ($NOEQ$) condition.

These conclusions were supported by the Type VI analysis of variance (see Lindquist, 1953, pp. 292-297). Treatments F was significant ($F = 3.09$, $p < .01$, $df = 6, 245$). Comparison of overall treatment means by t test yielded the following results: $SAMX:SBMX$, $t = 2.03$, $p < .05$, $SACT:SBCT$, $t = .65$, $p > .50$, $SANM:SBNM$, $t = 3.03$, $p < .01$. The t comparisons between each of the three SA treatments and the $NOEQ$ group were all significant beyond the .05 level.

The effect of the SA treatments tends to be more marked for materials from the second half of the experimental text than for the first. The facilitative effects which result from questions apparently grow with increased

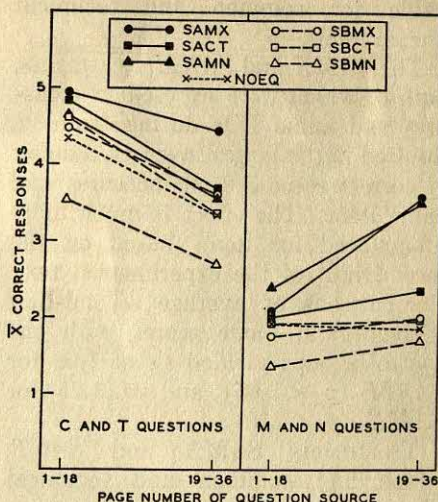


FIG. 1. Mean correct responses for criterion test questions of the two answer categories as function of location of the question source in the text.

exposure to the experimental questions at least for some items. This conclusion was supported by the significant interaction between Treatment $\times$ Location of criterion test question source in the text ($F = 2.70$, $p < .05$, $df = 6, 245$). This finding is consistent with the interpretation that EQ s in the SA -treatments tend to shape increasingly effective inspection behaviors.

Selective Facilitation Effects

Selective effects of experimental questions can be inferred where facilitation is more pronounced on criterion test items with answer types that have been represented in the set of experimental questions used in training. The results tend to support the selective facilitation hypothesis. This conclusion was reached in the following way:

Both the names:measures and common:technical phrases data were ordered in a manner consistent with the selective facilitation hypothesis although the effect was small and unre-

liable for common and technical phrase items.

The *SAMX* and *SANM* treatments, both of which were exposed to measures and name EQs during training, resulted in the largest average number of correct responses on measures and name items. The effect is much more pronounced for items based on the second half of the experimental text. Comparisons of average second-half measure and name scores with the control groups yielded *t*'s of 1.98 for *SANM* ($p < .05$) and of 1.75 for *SAMX* ($p < .10$).

Treatments *SAMX* and *SACT*, which had common and technical phrase items included in their experimental questions, resulted in a greater number of correct responses in the common and technical phrase category than the other treatments. This ordering was mainly due to technical phrase items. But the differences were so small that they were very likely due to change.

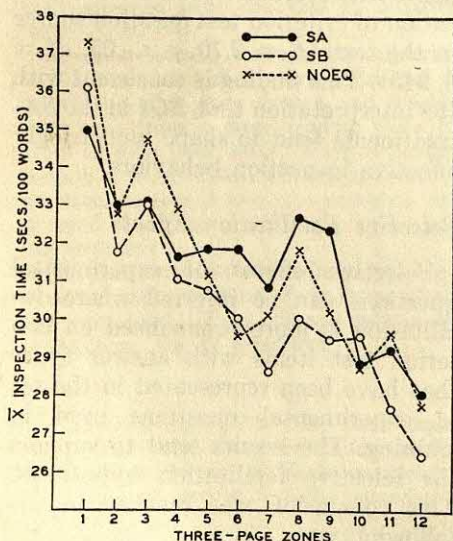


FIG. 2. Inspection time in average number of seconds per 100 words of text for the 12 three-page zones. (For the purpose of graphic simplicity, the data for the three *SA* treatments was pooled. So was the data for the three *SB* conditions.)

Inspection Time

Analysis of variance indicates no reliable differences in mean inspection time per page among the various treatments ($F = .80$, $df = 6$, 245). As can be seen in Figure 2, inspection time per page drops markedly as a function of the number of pages read. Average inspection time per page was significantly shorter for pages 19-36 than pages 1-18 ($F = 147.89$, $p < .001$, $df = 1$, 6). This finding is in keeping with previous observations (Rothkopf, 1966), and suggests an extinction-like process which can progressively weaken inspection behavior during the course of study.

DISCUSSION

This experiment provides further evidence that facilitative effects can be produced by questions that are presented immediately *after* inspecting the appropriate text segment (*SA* questions). Questions presented *before* inspection of the source segment (*SB* questions) were *not* found to produce these results.

For reasons stated in the introduction, the facilitative effects which were observed in this experiment have been identified with a class of activities called inspection behaviors or mathemagenic behaviors (Rothkopf, 1963, 1965). The results provide support for the hypothesis that the use of *SA*-type questions can modify mathemagenic behaviors so as to facilitate learning of specific subsets of the written experimental materials. A study by Hershberger (1964a) was equivocal about selective facilitative effects of questions.

The present experiment is not decisive as to the exact mechanism by which mathemagenic behaviors are modified by *SA* questions. The most plausible conception is a kind of adaptive evolution of mathemagenic be-

haviors, with questions providing selective contingencies. Mathemagenic behaviors, according to this view, are extinguished and dropped if they do not result in learning the skills necessary to answer the experimental questions. On the other hand, mathemagenic behaviors which preceded successful performance on experiment questions would be strengthened. As a consequence, the SA treatments would be more likely to result in the acquisition of adaptive (i.e., successful) mathemagenic responses than the SB or the NOEQ conditions. This conception would also explain why the effect of the SA treatments tend to be more pronounced for the second half of the experimental text, at least for some items. Increased exposure to experimental questions in the SA treatments increase the likelihood that successful mathemagenic behaviors have evolved.

It should be noted that the two categories of questions used in this experiment (i.e., names and measures versus common and technical phrases) are not equivalent. Questions requiring names, measures, or technical phrases for responses can be derived from only a limited number of sentences in the experimental passage while a question requiring common words as responses can be formed from nearly every sentence. The notion of selective facilitation presupposes the possibility of specifically adaptive mathemagenic maneuvers and this is quite difficult to conceive for the common word-type item. Consequently, it was not unexpected that facilitation effects for questions requiring common or technical phrase answers were found to be weak and were almost entirely due to the technical phrase questions.

The present experiment indicates that test-like events such as questions can facilitate learning. However, it

must not be overlooked that Rothkopf and Coke (1963, 1966) have shown that, under certain conditions, test-like events may shape mathemagenic behaviors which are contrary to instructional objectives. It is therefore important to avoid the conclusion that the effect of test-like events is always educationally desirable.

REFERENCES

- HERSHBERGER, W. A. Distinguishing errors of memory from errors of understanding by means of self-instructional tests. (Air-C28-8/64-TR) Palo Alto, Calif.: American Institute for Research, August 1964. (a)
- HERSHBERGER, W. A. Self-evaluational responding and typographical cueing: Techniques for programing self-instructional reading materials. *Journal of Educational Psychology*, 1964, **55**, 288-296. (b)
- KEISLAR, E. R. A descriptive approach to classroom motivation. *Journal of Teaching Education*, 1960, **11**, 310-315.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.
- ROTHKOPF, E. Z. Some conjectures about inspection behavior in learning from written sentences and the response mode problem in programmed self-instruction. *Journal of Programed Instruction*, 1963, **2**, 31-46.
- ROTHKOPF, E. Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.), *Learning and the educational process*. Chicago: Rand McNally, 1965. Pp. 193-221.
- ROTHKOPF, E. Z. Learning from written materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, 1966, **3**, 241-249.
- ROTHKOPF, E. Z., & COKE, E. U. Repetition interval and rehearsal method in learning equivalences from written sentences. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 406-416.
- ROTHKOPF, E. Z., & COKE, E. U. Variations in phrasing, repetition interval, and the recall of sentence material. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 86-91.

(Received April 7, 1966)

VARIATION IN THE AMOUNT OF IRRELEVANT CUES IN TRAINING AND TEST CONDITIONS AND THE EFFECT UPON TRANSFER¹

ROBERT L. R. OVERING

AND

ROBERT M. W. TRAVERS

McGill University

Western Michigan University

The principle of refraction was taught to 96 Ss (52 males, 44 females) under 1 of 2 conditions: (a) where a considerable amount of irrelevant visual information accompanied the instruction (TC+), and (b) where as much irrelevant visual information as possible had been eliminated (TC-). $\frac{1}{2}$ of each group was then tested for transfer under 1 of 2 test conditions: (a) in the presence of considerable irrelevant information; (b) in the absence of all possible irrelevant information. Ss trained under the TC+ condition showed no difference between test conditions. However, Ss trained in the TC- condition showed significantly better performance on the test in which irrelevant information had been reduced than on the test possessing irrelevant information.

An experiment previously reported by Overing and Travers (1966), building on the work of Judd (1908) and Hendrickson and Schroeder (1941), indicated that the *manner* in which a principle is taught affects the amount of transfer in a task requiring application of that principle. The task involved the application of the principle of refraction to the understanding of the bending of a beam of light at the surface of water. The results suggested that a "realistic" treatment with all its irrelevant cues was significantly better in facilitating transfer than was a "visually compressed" treatment which had many of its irrelevant cues removed. The designation of cues as relevant or irrelevant is a somewhat arbitrary matter. The criterion used here was subjective and, hopefully, logical. Those elements deemed to be essential (e.g., light beam and water) were left untouched, while those elements deemed to be irrelevant (e.g., shape of tank, ancillary

equipment) were eliminated or obscured.

In the previous experiment (hereafter referred to as Experiment I), the realistic treatment consisted of an actual demonstration of (a) the refraction of light rays entering and leaving water, and (b) the apparent displacement of an underwater object, when viewed at an angle from above. For the present demonstration a three-dimensional physical apparatus was used, and the verbal components of the instruction were transmitted by prerecorded tape. For the *visually compressed* treatment the same tape was used, thus ensuring that the verbal information was constant for each treatment, but the visual components of the demonstration of refraction and displacement were taught using life-sized, four-color line-drawings, designed to exclude all information or cues judged to be irrelevant.

One interpretation of the results of the previous experiment is that training with compressed information fails to give the learner experience in dealing with situations involving many irrelevant cues. This interpretation suggests that the person who learns to understand a principle in the absence

¹This article is based on research sponsored by the New Educational Media Branch, United States Office of Education, under Project C-997. Complete details of the procedures described can be found in a doctoral dissertation by R. L. R. Overing which includes this and other studies.

of irrelevant cues has more difficulty in applying the principle to a situation involving many irrelevant cues than he would have had if the original learning had been undertaken in the presence of irrelevant cues. The suggestion is that the advantage of learning under realistic conditions is that such conditions generally involve many irrelevant cues. Simplified situations, such as are presented by diagrams, do not involve irrelevant cues and, hence, may provide inefficient teaching situations. The study which follows represents an attempt to investigate further the extent to which learning either with or without irrelevant cues facilitates transfer to new situations which also may be either with or without irrelevant cues.

METHOD

Design and Materials

The experiment involved two $1 \times 2 \times 2 \times 2 \times 12$ factorial designs, each comprising one training condition, two sexes, two test conditions, two tests within each test condition and twelve repeated measures within each of the two tests.

Training Conditions

Of the two training conditions, one was identical to the realistic condition of the earlier experiment (i.e., the demonstration of refraction and displacement was conducted with actual physical apparatus and no effort was made to control for the presence of irrelevant cues) and for purposes of this experiment is designated here as Training Condition Positive (TC+). The second training condition, designated Training Condition Negative (TC-), used identical apparatus to TC+ except that every effort was made to remove or minimize irrelevant cues. To this end large black screens were built to screen off much of the apparatus and to provide a neutral background. Sections of the apparatus were coated with flat black paint to eliminate reflections and to minimize edges and surfaces of the apparatus. All ancillary apparatus such as light source, water pail, dust-making apparatus (used to make light beam show up in air) were all partially or completely

screened from the Ss. The S viewing the demonstration in this training condition could see little except the light beam passing through the air and water.

In both the negative and positive training conditions, the identical sound track transmitted the verbal component of the instruction, this sound track being the same one which had been used for the realistic treatment of Experiment I.

Testing Conditions

After training, Ss were exposed first to one of two test conditions. This first test, regardless of condition, is referred to as Test 1. They were then exposed to one of two conditions in a second test, referred to as Test 2. Both Tests 1 and 2 involved the problem of aiming a pistol correctly at an underwater target which, because of the refraction of light, did not really lie where it appeared to be.

Of the two test conditions, the one designated Test Condition Positive (Test +) was very similar to the test conditions of the original experiment. It was, however, somewhat embellished in that realistic pistols were substituted for the former plain wooden guns, new target objects suggestive of real organisms (a fish and a tadpole) shown in Figure 1 replaced the former simple cross and circle, and the numbers on the plastic grid (used to permit Ss to call off their point of aim) were represented both in red and blue, as opposed to being merely blue in the original experiment. Thus a number of new irrelevant conditions were introduced into this training condition.

In the other test condition, designated as Test Condition Negative (Test-), Ss were exposed to several modifications designed to eliminate or minimize all possible irrele-

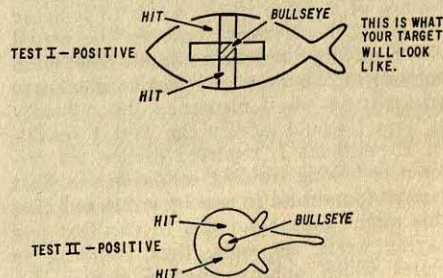


FIG. 1. Targets used in Tests 1 and 2 of Test Condition Positive as they appeared on the mimeographed sheets handed to S and as they appeared on the bottom of the tank.

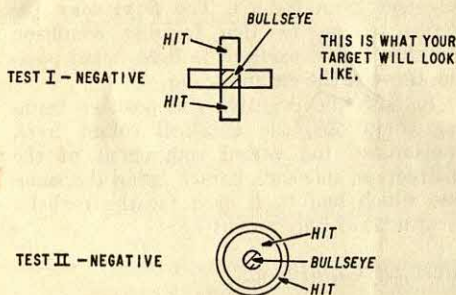


FIG. 2. Targets used in Tests 1 and 2 of Test Condition Negative as they appeared on the mimeographed sheets handed to S and as they appeared on the bottom of the tank.

vant cues. First, the tests of the negative condition were conducted in darkened rooms, so that as little as possible of the general surroundings were visible; second, the size and shape of the test tanks were obscured by covering the tanks with black cloth, in which was cut an aperture just large enough to permit the numbered portion of the grid to show. The target pistols were the simple wooden pistols of the first experiment, but were now covered with flat black paint. The target objects, as before, were the simple cross and circle shown in Figure 2. The only source of light was from a lamp placed under the black cloth so as to shine directly into the tank. Some light, naturally, escaped through the grid aperture and dimly illuminated the surroundings.

As in the previous experiment, each test condition comprised two tests, designated Test 1 and Test 2. However, the difficulty of the two tests in this second experiment was increased by changing the water depth in the tanks and altering the angle of incidence of the gun to the water to increase the visual displacement between the actual target and the apparent target. The main purpose of this increased displacement was primarily to see if increasing the difficulty of Test 1 would modify the Test 1 results. In Experiment I, reported earlier, no differences among training conditions on Test 1 were found and it was hypothesized that this might have been due to the first test having been too easy, thus resulting in scores which approximated a chance distribution.

Procedure

The Ss ($N = 96$) were sixth-grade pupils attending two elementary schools in a mid-

dle-class section of Salt Lake City. The group comprised 52 boys and 44 girls. As in the first experiment, Ss were preselected in the sense that those with IQ's below 100 or over 130 on the Pintner Intermediate Test were excluded. Pupils were randomly assigned to each experimental condition.

The procedure followed was very similar to that of the first experiment. Each of the four groups, each group comprising 13 males and 11 females, reported to the teaching station where they were exposed either to the negative training condition or to the positive training condition. Upon completion of the teaching program the group was then divided and one-half of the Ss reported to the Negative Test Station and the other half reported to the Positive Test Station. Ss entered the Test 1 area of their respective stations individually, and were given a printed sheet of directions which read as follows.

You have just learned that underwater objects, when viewed at an angle from above, do not really lie in the exact position that they appear to lie.

Remembering this fact, your problem, now, is to aim the pistol so that a bullet from it would strike the underwater target. The depth of the water over this particular target is $5\frac{3}{4}$ inches.

The plastic grid over the target has numbered lines marked on it. These lines will enable you to report your point of aim to the attendant.

When you have reported your point of aim, the examiner will say one of the following: "HIT," "MISS," or "BULLSEYE."

"HIT" means that you have hit some part of the target, either above or below the center of it.

"MISS" means that you have not struck any part of the target.

"BULLSEYE" means that you have struck the very center of the target.

TABLE 1
MEAN IQ OF ALL SUBJECTS BY TREATMENT

Treatment	$\bar{X}$ IQ	SD
Positive training—positive test	112.58	7.89
Positive training—negative test	115.12	8.32
Negative training—positive test	115.92	10.23
Negative training—negative test	114.21	9.49

Note.— $N = 13$ males, 11 females for each treatment. Total $N = 52$ males, 44 females.

Your object is to score a bullseye in as few shots as possible. You are to go on shooting and reporting your shots as *quickly* as possible until the attendant tells you to stop.

Any questions?

The instruction sheets to Ss at the negative and positive stations were identical except that in each case the appropriate target was drawn on them. The Ss were invited to question the attendant about the instructions if they were in doubt on any point. The Ss then aimed the gun so as to attempt to hit the underwater target, indicating their point of aim by reference to the numbered plastic grid, and the attendant provided feedback by saying "Hit," "Miss," or "Bullseye," whichever was appropriate, after each shot. The Ss, as before, were permitted 12 trials in which to achieve criterion. If they did so before the twelfth trial, they proceeded to the Test 2 cubicle where they were presented with the following instructions.

Your task here is similar to the one in the other room. You are to aim the pistol so that a bullet from it would strike the underwater target. The depth of the water over this target is $8\frac{1}{2}$ inches, that is, it is deeper than in the previous task. As before, you will be told "HIT," "MISS," or "BULLSEYE."

Any questions?

As with Test 1, the instructions for negative and positive conditions were identical except for the appropriate target representations.

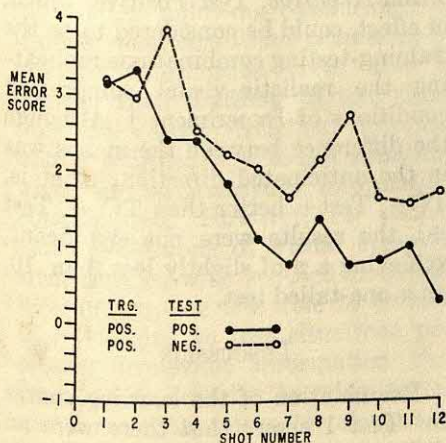


FIG. 3. Learning curves for Test 1, Training Condition Positive.

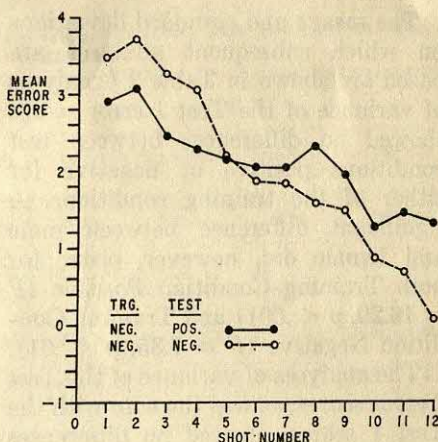


FIG. 4. Learning curves for Test 1, Training Condition Negative.

RESULTS

As in Experiment I, each S's scores on Tests 1 and 2 were converted to error scores by subtracting the value of the criterion aiming point from the value of the point of aim for each shot.

Figure 3 shows the learning curves which resulted from Test 1, Training Condition Positive, and Figure 4 shows the Test 1 learning curves resulting from Training Condition Negative. Figures 5 and 6 show the Test 2 learning curves for Training Conditions Positive and Negative, respectively.

Examinations of the individual learning curves indicated that, as in Experiment I, a majority of Ss had achieved criterion by Shot No. 7, and that the learning curves from Shot No. 7 on tended to be weighted heavily by the scores of one or two Ss who had shown little or no learning, and who in some cases, by this stage, were resorting to wild guessing. On this basis the same arbitrary decision was made as was made in the earlier experiment, to run the analyses of variance on the scores of the first seven shots only.

The means and standard deviations on which subsequent analyses are based are shown in Table 2. Analyses of variance of the Test 1 error scores showed no differences between test conditions positive or negative for either of the training conditions. A significant difference between male and female did, however, occur for both Training Condition Positive ($F = 16.20, p < .001$) and Training Condition Negative ($F = 7.35, p < .01$).

The analyses of variance of the Test 2 error scores, as was the case with the Test 1 scores, showed no differences between positive and negative testing conditions when learning had occurred under Training Condition Positive. A male-female difference was also evident, $p < .01$ ($F = 8.45, df = 1$ and 44) with this positive training condition. The results of the analysis of the Test 2 scores from the *negative*, or visually compressed, training condition are in sharp contrast, however. Here, a difference, significant at the .06 level ($F = 3.85, df = 1$ and 44),

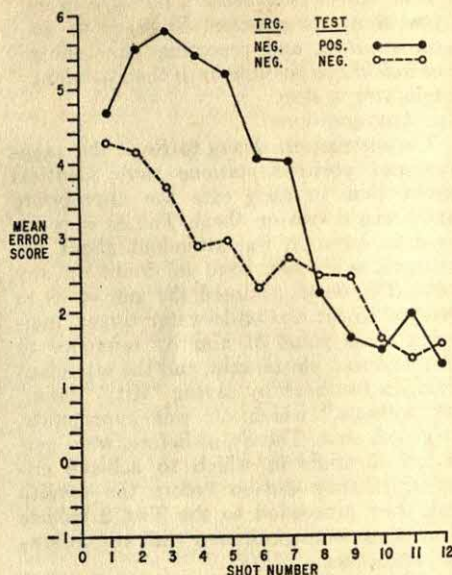


Fig. 6. Learning curves for Test 2, Training Condition Negative.

did occur between positive and negative testing conditions, and no male-female differences were evident.

In order to have a basis for comparing the results of Experiment I with those of Experiment II a t test was run between Training Condition *Positive*, Test Positive and Training Condition *Negative*, Test Positive, which, in effect, could be considered to be the training-testing combinations replicating the realistic-visual compressed conditions of Experiment I. Although the difference between the means was in the anticipated direction, (that is, TC+, Test+ better than TC-, Test+) the results were not significant, achieving a p of slightly less than .10, on a one-tailed test.

DISCUSSION

Examination of the learning curves for Test 1 shows that there were no differences in performance between those Ss trained under the condition with irrelevant information (TC+)

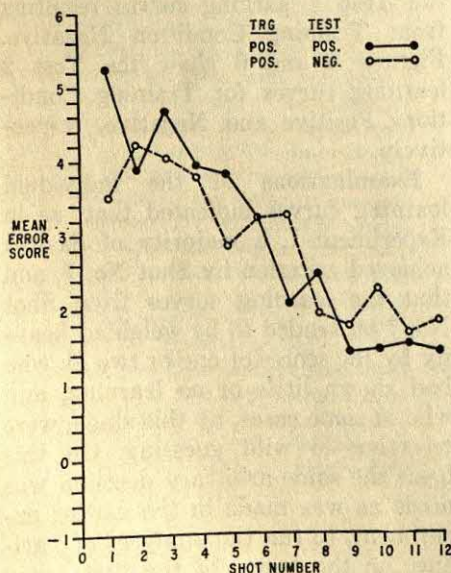


Fig. 5. Learning curves for Test 2, Training Condition Positive.

TABLE 2

MEANS AND STANDARD DEVIATIONS BY LEARNING CONDITION, TESTING CONDITION, SEX, AND TEST 1 AND TEST 2

Training condition	Test 1								Test 2							
	Irrelevant cues added (Test +)				Irrelevant cues reduced (Test -)				Irrelevant cues added (Test +)				Irrelevant cues reduced (Test -)			
	Male		Female		Male		Female		Male		Female		Male		Female	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Positive training	1.54	1.32	2.87	1.87	1.57	1.16	3.89	1.88	3.55	2.48	4.19	2.42	1.71	2.35	5.73	3.69
Negative training	2.03	1.46	2.93	1.57	2.17	1.23	3.51	1.45	4.59	3.08	5.40	3.53	2.84	2.71	3.80	2.43

and those Ss trained under the condition with reduced irrelevant information (TC-). This held true regardless of whether Ss were tested under the test condition with irrelevant information (Test+) or the test condition with reduced irrelevant information (Test-). This finding confirms the results of Experiment I where no differences between conditions were observed on Test 1.

However, as was also true in Experiment I, differences between certain of the training-testing combinations of Experiment II emerge in Test 2. These Test 2 results would seem to suggest the following conclusions: Ss trained under conditions possessing irrelevant information seem able to transfer their learning to test situations possessing irrelevant information, or alternatively, to test situations with reduced irrelevant information, *equally well*. However, Ss trained under conditions of visual compression, that is, with reduced irrelevant information, are less able to transfer their learning to test situations possessing irrelevant information than they are to transfer learning to test situations with reduced irrelevant information.

Although the results of this experiment do not appear to confirm the

findings of Experiment I, namely, that a TC+, Test+ combination is significantly better than a TC-, Test+ combination, the reader should note that both the Training Condition Negative and the Test Condition Positive of Experiment II were different from Experiment I. However, the significant difference between the TC-, Test- combination and the TC-, Test+ combination do appear to add support to the original supposition drawn from Experiment I that when irrelevant cues are encountered in the testing situation which were *not* present during training, the level of performance will be lower.

The male-female differences evident in the Test 1 results of both positive and negative training conditions, as well as in the Test 2 results of the positive training condition, but absent in the negative training condition, Test 2, require some comment. There is some evidence (Reed, 1961) that adolescent boys show more of the socially expected male interest in science and that they also report significantly more scientific activities than do girls. The subject matter taught in this present experiment is of a scientific nature and in addition the particular transfer task used appears to be by its nature a "masculine" task com-

prising activities and apparatus more familiar to boys than to girls. Girls, accordingly, could be presumed to be at a disadvantage which was reflected in generally poorer performance. Why this difference between male and female performance should disappear in Test 2 of Training Condition Negative is a phenomenon for which we have no explanation. Speculation might suggest that confronted with a learning task that was relatively unfamiliar and of low interest, females were able to learn more from the less distractive training condition in which irrelevant information had been reduced, and thus were able to perform comparably to males. What is of interest and suggestive of further research is the fact that having seen this phenomenon in Experiment II the authors returned to the data of Experiment I and found

a somewhat similar finding. That is, in Test 2 of Experiment I, females trained under the condition in which irrelevant information had been reduced performed comparably to males.

REFERENCES

- HENDRICKSON, G., & SCHROEDER, W. H. Transfer of training in learning to hit a submerged target. *Journal of Educational Psychology*, 1941, **32**, 205-213.
- JUDD, C. H. The relation of special training to general intelligence. *Educational Review*, 1908, **36**, 28-42.
- OVERING, R. L. R., & TRAVERS, R. M. W. Effect upon transfer of variations in training conditions. *Journal of Educational Psychology*, 1966, **57**, 179-188.
- REED, H. B., JR. Teacher variables of warmth, demand and utilization of intrinsic motivation related to pupils' science interest: A study illustrating several potentials of variance-covariance. *Journal of Experimental Education*, 1961, **29**, 205-229.

(Received April 14, 1966)

Help him get ready for work

The new Rochester Occupational Reading Series offers both reading and language instruction, plus occupational information to poor readers who will soon be entering the job world.

By reading realistic stories and completing exercises that are mature in approach, learners gain a clear understanding of the skills and attitudes necessary for job success. They are led to develop such skills as completing application forms, reading comprehensively, performing simple arithmetical computations, and speaking simply and correctly. Basic social attitudes include: good manners, obedience to law, and good health practices.

To use the program, a learner only needs second through fifth grade reading ability. All materials are written at three different levels within that range. They appear to be identical so older students working at the lowest level feel no embarrassment about their status.

Prepare your students for the occupational world. Write for a brochure describing this unique learning program — the new Rochester Occupational Reading Series.

S R A

Science Research Associates, Inc.
259 E. Erie St., Chicago, Ill. 60611

A Subsidiary of IBM





EDUCATIONAL PSYCHOLOGY

An Application of Social and Behavioral Theory
by LOUIS M. SMITH and BRYCE B. HUDGINS

both of Washington University at St. Louis

Treating the principles and substantive findings of educational psychology in a manner that is at once scholarly, practical, and thoroughly sound, Professors Smith and Hudgins present an integrated view of social and learning behavior in the classroom.

"An excellent text! I am particularly appreciative of the incorporation of material from social and behavioral theory."

Sidney Yudin, *New York University*

Knopf, 576 pages; 21 figures; 73 tables; \$8.50

An INSTRUCTOR'S MANUAL, is available.

SELECTED PAPERBOUND BOOKS

In preparation . . .

SOCIETY'S CHILDREN

A Study of Ressentiment in Secondary School Education

by CARL NORDSTROM, *Brooklyn College*, EDGAR Z. FRIEDENBERG, *University of California, Davis*, and HILARY A. GOLD, *Brooklyn College*

Spring 1967; SED7; 224 pages; \$2.45

GUIDING THE GIFTED CHILD

by GEORGE I. THOMAS, *State Department of Education, Albany, New York*, and JOSEPH CRESCIMBENI, *Jacksonville University*

1966; SED4; 218 pages; \$1.95

LEARNING Reinforcement Theory

by FRED S. KELLER, *Columbia University*

PP13; 47 pages; \$.95

Just published . . .

COMING OF AGE IN AMERICA

Growth and Acquiescence

by EDGAR Z. FRIEDENBERG

February 1967 V368 316 pages; \$1.95

A truly revolutionary critique of American education which documents the clash between the adolescent and the values of the upwardly-mobile mass culture purveyed by our school systems.

Recently published . . .

LANGUAGE AND THE DISCOVERY OF REALITY

A Developmental Psychology of Cognition

by JOSEPH CHURCH, *Brooklyn College of The City University of New York*

V338 \$1.95

For further information, write to

RANDOM HOUSE & ALFRED A. KNOPF, Inc.

College Department

501 Madison Avenue

New York 10022

FOR CURRENT RESEARCH IN CHILD PSYCHOLOGY

CHILD DEVELOPMENT: An Individual Longitudinal Approach

LELAND H. STOTT, *Leader of the Longitudinal Research Program,
The Merrill-Palmer Institute, Detroit, Michigan*

Using a multidisciplinary approach, this book illustrates the basic principles of human development and behavior. Using a case history approach, the author traces the development of two individuals, a pair of twins (boy-girl), and applies basic principles to data from their files. The major theme of the book is that both the development of a physical structure and changes in the functioning of that structure come about as a result of the interaction between the organism (the individual as he is) and his unique environment.

March 1967

448 pp.

\$6.95 tent.

CHILDREN: Behavior and Development, Second Edition

BOYD McCANDLESS, *Emory University*

A substantially rewritten version of the author's earlier work, *Children and Adolescents*, (1961) which omits, except for the material on physical growth, the earlier book's treatment of adolescents. McCandless stresses the general principles that are helpful in understanding children as they function in society, with special attention given to the child's life in school. Since the publication of his earlier book, the author has modified his theoretical framework to recognize the impact of social learning and its implications for school children, especially where Piaget's theories are concerned.

May 1967

640 pp.

\$8.50 tent.



**Holt, Rinehart
and Winston, Inc.**

383 Madison Avenue,
New York, New York 10017

Recently Published—

The Causes of Behavior

READINGS IN CHILD DEVELOPMENT AND EDUCATIONAL PSYCHOLOGY, SECOND EDITION

Edited by **Judy F. Rosenblith**, Wheaton College; and **Wesley Allinsmith**, University of Cincinnati.

Used by hundreds of colleges and universities in the first edition, this eminently successful book has been revised from suggestions made by its users. Approximately 40% of the articles are new. Of these new selections, approximately one-half have appeared since 1961, many as recently as 1965. The second edition features an expanded discussion of the culturally disadvantaged child, and a more thorough discussion of Piaget. A test manual is available.

1966. 8½ x 11. Paperbound. 608 pp. \$6.50

Recently Published—

An Introduction to Statistics and Measurement

A PROGRAMMED BOOK

By **Lowell A. Schoer**, State University of Iowa

The text presents statistical concepts necessary for dealing with norms, predictions, validity, reliability, regression, and the standard error of measurement.

"Mr. Schoer has done an excellent job of covering the material needed in a first course in Statistics and Measurement. His arrangement of frames is in easily understood [form], and his instructions are sufficient and clearly stated." —Dr. Earl D. Ludlow, Missouri State Teachers College

1966. 8½ x 11. Paperbound. 201 pp. \$3.95

Widely used—

Educational Psychology, Second Edition

By **James M. Sawrey** and **Charles W. Telford**, both of San Jose State College

A comprehensive presentation of the psychological foundations of education. The authors stress the principles of motivation and learning as the central concepts of educational psychology.

1964. 6 x 9. 592 pp. \$7.95

A Student Guide by **Earl L. Jandron** of San Jose State College accompanies the text. Each chapter includes class-tested programmed units. 1964. 8½ x 11. Paperbound. 184 pp. \$3.75.

Dept. Z-1, 150 Tremont St., Boston, Mass. 02111

Coming Spring, 1967—

Evaluating Pupil Growth:

PRINCIPLES OF TESTS AND MEASUREMENT, THIRD EDITION

By **J. Stanley Ahmann**, Academic Vice President and Professor of Psychology, Colorado State University; and **Marvin D. Glock**, Professor of Educational Psychology, School of Education, Cornell University.

This revision of a widely acclaimed book is designed for courses in Tests and Measurements in the departments of Education, Educational Psychology, General Psychology, and Physical Education.

Thoroughly modern, the text:

Contains up-dated discussions of standardized instruments; for example, the Differential Aptitude Test.

Presents specific examples of measuring educational and behavioral objectives for considering the use of computers in text scoring, analysis, and reporting.

Includes new discussion questions and recent suggested readings.

1967. 6 x 9. Est. 500 pp. \$7.95 tent.

ALLYN AND BACON, INC.

What Psychologists Say About the Latest MMY

THE SIXTH MENTAL MEASUREMENTS YEARBOOK

Edited by Oscar Krisen Buros

► "most recent and definitive consumers' report of the psychometric world. This edition for the first time lists all the tests that are currently in print.* The 396 contributing authorities read like a Who's Who in the world of testing.* the *Mental Measurements Yearbooks* certainly are the test user's best defense against inadequate tests and hard selling test salesmen. Every counselor who makes use of tests and every professional person who is involved with building testing programs should have this book available to him.* If you have been using the *Mental Measurements Yearbooks* before, nothing needs to be said except, 'He has done it again.' If you are not familiar with this book, insist to your principal or supervisor that it must be available for your use."—BUFORD STEFFLRE, *Personnel & Guidance Journal*, April 1966

► "*The Sixth Mental Measurements Yearbook* is a 1,714 page collection of reviews of tests and books which is as indispensable to personnel psychologists as Webster's dictionary, Roget's thesaurus, and the APA directory. There can be no question about the usefulness of the volume.* In 8 of the 14 reviews of the Fifth Yearbook (which—with characteristic completeness—Professor Buros has included in the Books and Review section of the current Yearbook) the adjective *monumental* is used in reference to the Yearbook. This continues to be highly appropriate." — THOMAS L. HILTON, *Personnel Psychology*, Autumn 1966

► "An evaluation of this last volume can be couched, with palpable redundancy, in terms applied by the reviewers of previous volumes in the series: This remarkable, monumental, colossal, and elaborate book, representing gargantuan and heroic efforts applied compulsively and with missionary zeal is certain to become a 'must'—perhaps even the indispensable 'bible'—for the increasing number of test users.* In relation to general price increases the Yearbooks are still 'best buys'."—DOROTHY C. ADKINS, *Science*, March 18, 1966

► "An editorial masterpiece and a beautifully printed volume which is a tribute to the assiduous efforts expended by Oscar Buros over a period of nearly six years, *The Sixth Mental Measurements Yearbook* is, as one would expect from the Buros tradition, the best one to date."—JOAN J. BJELKE and WILLIAM B. MICHAEL, *Educational & Psychological Measurements*, Summer 1966

► "For over a quarter of a century psychologists have relied heavily upon the *Mental Measurements Yearbooks* for information about tests they wish to use. This new volume will immediately and inevitably find its place alongside the earlier ones. These do not, of course, become obsolete, but are supplemented by the latest volume and with it and the general index, *Tests in Print* (1961) by the same editor, form a unique and comprehensive body of knowledge about all known tests published in English.* tremendous service to psychometry."—L. B. BIRCH, *British Journal of Educational Psychology*, June 1966

Published November 1965, 1752 2-column pages, \$32.50

The Gryphon Press, 220 Montgomery Street,

Highland Park, New Jersey 08904

Three From Macmillan

ESSENTIALS OF LEARNING Second Edition

By Robert M. W. Travers, Western Michigan University

This highly readable and thoroughly updated edition provides unusually broad coverage of the field of learning, and discusses individual topics within the field in specific detail. Thus, not one theory of learning is presented, but many, each firmly rooted in research, providing the student with significant and current data on every aspect of learning. The author analyzes the possible implications that modern studies of learning may have for curriculum-planning, for teaching methods, and for general educational planning.

1967, approx. 544 pages, \$7.95

EDUCATIONAL PSYCHOLOGY Second Edition

**By Glenn Myers Blair, R. Stewart Jones, and Ray H. Simpson,
all of University of Illinois**

The second edition of this distinguished text offers a functional and unusually comprehensive treatment of educational psychology. It carefully illustrates psychological theories with actual classroom examples and case studies. The material includes teaching machines, feedback, "brainstorming," creative thinking, and discovery and learning.

1962, 678 pages, \$7.95

CLASSROOM GROUP BEHAVIOR

**By Mary A. Bany, and Lois V. Johnson, both of California State
College, Los Angeles**

This is the first textbook to provide teachers with an understanding of the factors which influence group behavior in the classroom. The authors provide significant insights into the effects of the group upon the individual, and they suggest practical techniques for the control and guidance of small groups. Study questions accompany each chapter.

1964, 464 pages, \$6.95

Write to the faculty service desk for examination copies

THE MACMILLAN COMPANY
866 Third Avenue, New York, N. Y. 10022

NEW FROM SCRIBNERS

The Psychology of Childhood: A Survey of Development and Socialization

Karl G. Garrison and Albert J. Kingston
In Association with Harold W. Bernard

This comprehensive new text has been carefully designed to provide a factual, well-documented account of the many facets of the child and his psychological development. The book takes a biosocial approach to its subject, emphasizing the importance of family, school, and peers in the child's character and personality growth, as well as in his social development. Particular attention is paid to the early foundations of development in the pre-school and earliest school years. There is an annotated bibliography, an appendix listing pertinent films, subject and name indices, and numerous figures, tables, and photographs. A paperbound workbook has been prepared by Professors Garrison and Kingston and by Professor Ocie Dekle of the University of Georgia. An *Answer Key* is available to teachers.

1967. xxviii, 457 pages. \$7.25



CHARLES SCRIBNER'S SONS

COLLEGE DEPARTMENT

597 Fifth Avenue • New York 10017

Publication Manual of the American Psychological Association

1957

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. Scientists writing for publication will find the Publication Manual an indispensable guide.

Price \$1.00

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION

1200 Seventeenth Street, N.W., Washington, D. C. 20036

EDUCATIONAL PSYCHOLOGY

Second Edition

by *LEE J. CRONBACH, Stanford University*

A leading textbook in its field, *Educational Psychology* provides an integrated treatment of cognitive development and a full presentation of the psychological theory and research that underlie modern educational practice. 706 pages, \$8.50.

Accompanying the text:

Student Guide by Gabriel M. Della-Piana. Designed to help the teacher and prospective teacher apply the principles of educational psychology to concrete teaching situations.

Test Item File prepared by Jacob S. Orleans and Lee J. Cronbach. Three separate files are now available.

LEARNING

by *J. CHARLES JONES, Bucknell University*

To be published in March—a concise interpretation of recent psychological research findings on learning that provides the teacher-in-training with a theoretical basis for examining procedures and problems of education. The author first considers education and then proceeds to a discussion of important related topics. Brief, clearly written, and thoroughly readable, this text is especially suitable for Introduction to Education courses and for courses that emphasize the psychological foundations of education. *Learning* is in the Professional Education for Teachers Series.

Paperbound, 192 pages, \$1.95 (probable). Publication: March



Harcourt, Brace & World, Inc.
New York / Chicago / San Francisco / Atlanta

RUSSIAN MONOGRAPHS ON BRAIN AND BEHAVIOR

A series of translations of Russian Monographs concentrating upon the historical background of investigations in neurophysiology and conditioned behavior as well as upon certain contemporary advances in these fields.

The translation program was initiated by H. W. Magoun and D. B. Lindsley and has been continued by the latter. The complete series will consist of eleven monographs published in hard-cover book form. To date, three of the eleven monographs have been printed and are available through the American Psychological Association.

Available now . . .

ESSAYS ON THE HISTORY OF PHYSIOLOGY IN RUSSIA

By Kh. S. Koshtoyants.

These essays review broadly the history of the development of physiology as a science in Russia, beginning in the early 18th century and continuing to modern times. They are more than a history for they deal with the personal characteristics and strivings of great scientists who forged teaching and research disciplines in anatomy, physiology, and psychology.

321 pp.

\$4.00

AUTOBIOGRAPHICAL NOTES

By I. M. Sechenov.

Sechenov, revered as the father of physiology in Russia, was born in 1829. His scientific and academic pursuits, though always stimulating and challenging, were fraught with difficulties due to political interference and harassment. These **Autobiographical Notes**, written about a year before his death in November, 1905, provide a vivid picture of his history.

174 pp.

\$4.00

ORIENTING REFLEX AND EXPLORATORY BEHAVIOR

Edited by L. G. Voronin, A. N. Leontiev, A. R. Luria, E. N. Sokolov, and D. S. Vinogradova.

A collection of papers read at a Conference organized in February, 1957 by the Institute of Defectology of the Academy of Pedagogical Sciences, RSFSR, and by the Department of Psychology and Higher Nervous Activity of the M. V. Lomonosov Moscow State University. As the varied experiments and points of view in this collection of papers will reveal, a serious attempt has been made to establish the role and significance of the orienting-exploratory reaction, not only in the elementary behavior of lower animal forms, but also in the complex behavior of higher forms, including man and his cognitive functions.

462 pages

\$4.00

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1200 Seventeenth St., N. W., Washington, D. C., 20036

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 58

April 1967

Number 2

CONTENTS

<i>Replacement and Nonreplacement Strategies in Children's Problem Solving</i>	69
M. C. WITTRICK	
<i>Some Psychological Aspects of Subject-Matter Structure</i>	75
PAUL E. JOHNSON	
<i>Free-Operant Preference for One of Two Stories: A Methodological Note</i>	84
THOMAS C. LOVITT	
<i>Social Schema of Normal and Disturbed School Children</i>	88
RHODA LEE FISHER	
<i>Identification Patterns, Motivation, and School Achievement of Bright Junior High School Boys</i>	93
THOMAS A. RINGNESS	
<i>Effects of Strategy, Sex, and Age on Conceptual Behavior of Elementary School Children</i>	103
GLENN E. TAGATZ	
<i>Implicit Associative Response Occurrence in Learning with Retarded Subjects: A Supplementary Report</i>	110
WILLIAM P. WALLACE	
<i>Second-Order Ability Structure Revealed in Rights and Wrongs Scores</i>	115
JOHN L. HORN AND WILLIAM J. BRAMBLE	
<i>Erratum</i>	122
<i>Discussion versus Memory in Cooperative Group Concept Attainment</i>	123
PATRICK R. LAUGHLIN AND MARY AUSTIN DOHERTY	

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 82

April 1991

Number 1

CONTENTS

1. Assessment and Measurement Strategies in Gifted and Talented Education
M. D. WITTEN
60
2. Some Psychological Aspects of Gifted Talent
WILLIAM E. JENNISON
75
3. A Practical Procedure for One of Two Types of Psychological Tests
THOMAS A. JENNISON
84
4. Social Effects of Internal and External School Control
THOMAS A. JENNISON
92
5. Identification Patterns, Standards, and Social Adjustment of Gifted and Talented High School Boys
THOMAS A. JENNISON
99
6. Effects of Student Self and Peer Conceptualization of Homosexuality
THOMAS A. JENNISON
103
7. Student Attitudes and Responses to Learning with Technical Subjects
WILLIAM E. JENNISON
110
8. Second-Order Ability Measures Derived in High and Average Groups
JOHN E. JENNISON AND WILLIAM E. JENNISON
115
9. Discussion
122
10. Discussion: Student Attitudes in Education
WILLIAM E. JENNISON
125

REPLACEMENT AND NONREPLACEMENT STRATEGIES IN CHILDREN'S PROBLEM SOLVING

M. C. WITTRICK

University of California, Los Angeles

The replacement and nonreplacement strategies were operationally defined in problems involving 4 concepts. The 119 2nd graders from 2 public schools were individually assigned at random to the 3 treatments: Replacement, Nonreplacement, and Control. From a "hierarchical" model of verbally mediated transfer the following prediction was made. On tests of learning, near and remote transfer, from high to low the treatment means would rank: Nonreplacement, Replacement, and Control. In each of the 4 tests of learning and transfer, the treatment effect was statistically significant ($p < .01$) by analysis of variance. The predicted rank order among the means occurred in each school in each test. Newman-Keuls multiple comparisons tests indicated close agreement to the hypothesis. Young children learned a difficult strategy. They transferred it to new problems, including problems involving only unrehearsed concepts.

Experiments on the teaching of problem solving to children are rare in educational psychology. But these experiments can help educational psychologists to discover knowledge and theory about what is possible in teaching and in transfer of training.

The experiment reported below involves two types of problems. First, the dependent variable of most interest, transfer to new concepts, is often impervious to instruction, especially with young children. Second, an efficient nonreplacement strategy is difficult for young children to learn.

In two experiments on verbal cues (Wittrock & Keislar, 1965; Wittrock, Keislar, & Stern, 1964) children were taught to transfer to new examples of the same concepts learned during training. In the instructional treatment best for producing transfer within the previously learned concepts, the child was cued with the correct concept name. Neither the general basis for matching, nor the specific, correct answer to each item, which were the cues used in two other treatments, was as effective as the correct concept name. But none of the treatments produced much transfer to

new concepts. It was concluded that something more than these verbal cues, probably a verbal sequence or a strategy¹, is needed to produce transfer to new concepts. A strategy is general and should transfer broadly.

Anderson (1966) taught first graders to use a strategy to solve problems. However, he used as subjects (Ss) only the top third of the IQ distributions of the classes in the study.

From information theory, both the replacement and nonreplacement strategies used below should be more efficient than the "try one at a time" strategy Anderson used. In the study below, the child is taught to gain one bit of information on each succeeding trial. Until he solves the problem, on each trial he can eliminate half of the remaining choices.

When they are solving problems, children and adults often sample with replacement. That is, within a problem, they sometimes repeat incorrect answers and hypotheses. In their theoretical models of concept learning and problem solving, Restle (1962) and

¹ A strategy is operationally defined as a discriminable set of responses to a series of stimuli.

Atkinson, Bower, and Crothers (1966) assume that people often sample with replacement.

A more efficient procedure is to sample only from the population of answers or hypotheses that has survived elimination in previous trials. This procedure is called sampling without replacement. It should produce greater learning and transfer to problems involving a pattern of eliminating incorrect answers, even when new concepts replace the ones used during training. To build upon the studies by Anderson and by Wittrock et al. described above, a strategy of sampling without replacement was taught to the entire IQ distribution of two second-grade classes of public school children.

The hypothesis tested in this study is as follows. With young children solving problems involving four concepts, the nonreplacement strategy produces learning, near transfer, and remote transfer greater than does the replacement strategy, followed in turn by the control procedure. The nonreplacement strategy was operationally defined by having the children eliminate in sequence cards which named incorrect concepts. The replacement strategy was like the nonreplacement strategy, except that the child was not taught to eliminate the cards for the incorrect hypotheses.

The basis for the hypothesis is a "hierarchical" model of verbally mediated transfer. Verbally mediated transfer is hypothesized to be a function of the hierarchies of verbal associations the learner brings to the problem. These include his preferences for choosing among competing hierarchies associated to the problem, the dominance of the levels within each hierarchy, his habits for sequentially selecting among levels within a hierarchy, and his habits of sequentially

selecting responses within a given level of a hierarchy.

Word-association norms provide one estimate of superordinate, coordinate, and subordinate associations within hierarchies (cf. Miller, 1951). But, in this study, only children's strategies for choosing among responses all of one level of a hierarchy were studied. Therefore, predictions about superordinate and subordinate associations will not be discussed here. Sampling without replacement is an uncommon but efficient strategy for selecting hypotheses from one level of a hierarchy.

METHOD

Experimental Design

A fixed effects model, simple randomized design was used. The Ss were individually assigned at random to one of three treatments: Nonreplacement, Replacement, and Control. Each S was successively given familiarization training, preexperimental training, random assignment to one of three experimental conditions, experimental training, and the posttests.

Subjects

The Ss were 119 second grade students, 66 males and 53 females, the entire second grade populations from two Los Angeles public elementary schools.² From the original group of 128 children, nine were dropped because of frequent absence. The means and standard deviations of the treatment groups' MA, CA, and IQ, from the California Public School Primary Test of Mental Abilities, are presented in Table 1. These measures show no particularly divergent means or variances for any of the treatment groups.

² The author wishes to thank Sue Buckner who wrote the instructional materials and gathered the data. Thanks are due to the following people who made the study possible: Thomas Reece, Area Superintendent, Elementary Area West, Los Angeles City Schools; William Haley, Principal of the Marquez Elementary School; and Ann Louise Powell, Principal of the Richland Avenue Elementary School.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF THE
TREATMENT GROUPS' MA, CA, AND IQ

Treatment condition	MA	CA	IQ
Nonreplacement	96.9	84.9	114.1
	11.2	2.2	12.1
Replacement	98.4	85.6	115.4
	10.2	3.5	13.6
Control	94.7	85.6	111.2
	10.4	3.8	13.3

Apparatus

Each *S* was seated in one of 10 booths in a school room reserved for the experiment. He saw slides on a screen at the front of the room. The instructions were recorded on magnetic tape and played through ear-phones to each child. He responded to each problem by pushing a button on his two-choice response box. Each of the two buttons on the response box would automatically turn green or red to indicate that a choice was correct or incorrect, respectively.

In addition to the slide projector and the response boxes, the apparatus consisted of a master control panel, a stereo magnetic tape deck, and a paper tape data recorder. The slide projector was modified with photo-electric cells, which actuated the master control panel, displayed for the experimenter the correct choice for an item, and provided immediate feedback to each child. One channel of the stereo magnetic tape provided instructions to the children. The other channel of the tape provided impulses to control the automatic apparatus. Once the experimenter loaded the slide projector and started the tape, the presentation of stimuli, the control of the length of the response interval, the immediate feedback to *S*, and the recording of the responses were all automated.

Each *S* had four cards, each with words and designs to represent the four concepts: color, size, shape, and number. On the inside surface of the front vertical wall of his booth were hooks to hang the cards. The experimenter manually recorded the placement of cards on the hooks. The data recorder registered only the button presses.

Materials

Each slide presented a matching-to-sample stimulus which varied in four ways: color, size, shape, and number. The stimu-

lus to be matched was centered toward the top. The two choices for matching were below it, one on either side. Each problem involved a series of six or more slides. The basis for matching one of the two bottom pictures to the top picture was always one of the four concepts given above. The instances of each of these four concepts were as follows. For color, they were blue, green, yellow, and red. For shape they were cat, circle, diamond, and house. Large and small were used for size; one, two, three, and four were used for number.

The posttests measured learning and transfer. The tests of learning were made from the materials described above. The first transfer test introduced new colors (black, orange, purple, and green) and new shapes (boat, swan, square, and cone).

The remote transfer test introduced four new concepts. Triangles varied in the direction of their apexes (up or down), the texture of their interiors (dotted or plain), their rectangular borders (solid or dotted), and the position of a short horizontal line inside the border (at the base or at the apex of the triangle).

Procedure

With the exception of the experimental training, the procedure was identical for all groups. Each child worked individually in a booth. Children participated in groups of 10. A session lasted approximately 12 minutes. There were 13 days of training. The 4 days of testing were given immediately after training.

Familiarization training. In one session the children were given a series of eight problems. Each child was taught to push the button which corresponded to the one of the two bottom pictures identical to the top picture.

Preexperimental training. The object was to establish associations between the concepts and the cards to be used in experimental training. The first problems involved two concepts—color and shape—and two instances of each, green and red, and circle and cat. For each child, the task was to hang on the hook the card which illustrated the basis for matching the bottom picture to the model picture. After these problems, three instances of each of the two concepts were introduced. This gradual progression of difficulty was followed until all four concepts and all given instances of each had been learned. A test over training was then given. It consisted of eight prob-

lems of six slides each. The children who scored 80% or more correct on this test were assigned to the experimental training.

Experimental training. All 119 of the children who regularly attended the sessions successfully completed the preexperimental training and were individually assigned at random to one of the three treatments. The experimental training lasted four days and consisted of eight problems of six slides each. All problems and all slides were identical for all three treatment groups. The treatments differed only in the verbal instructions. The first slide of each problem was always designed to enable the child to eliminate half of the four choices. The lower left picture would always match the top picture in two ways while the lower right picture would always match the top picture in the other two possible ways.

The children in the Nonreplacement Group were taught to hang the four concept cards on the four hooks in front of them. They hung the cards to represent the ways the right and left bottom pictures matched the top picture. They then pressed buttons to find which of the two lower pictures was the correct match. Each child then took from the hooks and turned face down the two cards which represented the incorrect bases for matching. He used the two remaining cards on the next slide in the set. Again he pressed a button to find the correct match, and he turned face down the card which did not represent the correct basis for matching. The one remaining card, which should now be the correct choice, was used to match the other slides in the set.

With one exception, the Replacement Group was taught the procedure described above for the Nonreplacement Group. No mention was made of turning face down those cards eliminated when they were incorrect.

In the Control Group, Ss were given the procedure described above for the Replacement Group, but with one exception. They were given the cards without instruction to hang them underneath the right and left choices. They were instructed to find the one of four rules for matching. They were told to use the cards, if they wished, in any way they would like.

Testing. Four posttests were presented immediately after the experimental learning. These tests are described above in the Materials section. The first test, a test of learning, presented the same concepts and instances used during experimental training. The second test measured transfer to

new instances of the concepts color, and shape. Each child was allowed to use his four concept cards on these two tests.

The third test was identical to the second except that the children were not permitted to use their concept cards. The fourth test measured remote transfer. The concepts and instances of this test had never previously appeared in the experiment. No cards were permitted on this test.

At the beginning of each of the four tests, the children were told all of the four possibly correct bases for answering the problems. The child's task was to remember these bases and to determine the correct one for each problem. At no time was he given a cue or strategy during any test.

RESULTS AND DISCUSSION³

A heterogeneity of variance test (Edwards, 1960, p. 105) was run on each dependent variable. The assumption of homogeneity of variance was not rejected for any of the four variables. A Treatments $\times$ Schools analysis of variance, fixed effects model, was performed on each of the four dependent variables. The F ratio for the treatment mean square was statistically significant ($p < .01$) for each of the four dependent variables. The F ratios were 7.28 for learning, 9.47 for transfer to new instances with cards, 17.62 for transfer to new instances without cards, and 5.62 for transfer to new concepts. The schools effect was statistically significant ($F = 4.85, p < .05$) on the transfer-to-new-concepts test. There was no statistically significant interaction on any of the four dependent variables.

In Table 2 the means and standard deviations for the four dependent variables are listed by treatments within each of the two schools. From high to low, the predicted rank order, Non-

³The UCLA Campus Computing Facility and the UCLA Biomedical Computing Facility were used for the statistical analyses of the data. The research was financed by the Ford Foundation Fund for the Advancement of Education.

TABLE 2
MEANS AND STANDARD DEVIATIONS OF THE POSTTEST SCORES

Treatment groups	Learning (with cards) 40 items	Transfer to new instances (with cards) 40 items	Transfer to new instances (without cards) 40 items	Transfer to new concepts (without cards) 25 items
School I				
Nonreplacement group	29.4 4.4	29.6 5.2	30.1 4.0	17.8 3.5
Replacement group	25.1 6.3	24.9 7.1	26.4 7.6	16.9 5.5
Control group	22.4 5.2	20.6 4.7	20.9 6.5	14.2 3.6
School II				
Nonreplacement group	25.7 8.5	25.0 8.5	27.7 6.5	16.1 5.0
Replacement group	22.9 6.5	24.0 6.1	23.3 6.2	14.4 5.2
Control group	21.7 7.2	20.9 8.0	20.1 6.4	12.8 4.7

Note.— $N = 19$ for the control group in School I; for all other groups, $N = 20$.

replacement, Replacement, and Control, occurred on each of the four tests. This rank order occurred in each test in each school.

A Newman-Keuls multiple comparisons test (Winer, 1962, p. 80) was run on each of the four dependent variables. For the learning test, the Nonreplacement Group and the Replacement Group each had significantly higher means than the Control Group ($p < .05$). But there was no statistically significant difference ($p > .05$) between the Nonreplacement and the Replacement Groups. Both experimental groups learned better than did the Control Group.

On the transfer-to-new-instances test with cards, the mean of the Nonreplacement Group was statistically greater than the mean of the Replacement Group or the mean of the Control Group. The Replacement Group and the Control Group's means were not statistically significantly different from each other. The Nonreplacement Group transferred better than either of the other groups.

For the test of transfer-to-new-in-

stances without cards, all differences among the means of the three groups were statistically significant ($p < .05$). The Nonreplacement Group was highest, the Replacement Group next, and the Control Group last. The results of this test were as predicted.

On the last test, transfer-to-new-concepts without cards, the mean of the Nonreplacement Group was greater ($p < .05$) than either the Replacement Group or the Control Group. The Replacement and the Control Group means were not statistically significantly different from each other.

In brief, the results are in close agreement with the predictions, especially on the transfer-to-new-instances test without cards. The children in the experimental groups learned to use the strategies. They transferred the strategies to new problems. Transfer was best for the nonreplacement strategy.

The most exciting finding was that the nonreplacement strategy transferred to problems where the cards were removed. The absolute mean dif-

ferences and the statistical significance of the differences among the means increased on these two tests compared with the tests of learning and transfer-to-new-instances with cards. The children not only learned to use the strategies, but they could transfer them to problems without the cards or props, even to difficult problems involving transfer to four new concepts.

These results occurred in two public school systems using the entire second grade population of each school, not with a highly selected sample of children. The author's previous attempts at obtaining remote transfer with verbal cues were not effective. However, a nonreplacement strategy (call it a verbal sequence if you prefer) was effective in this study on the same transfer tests used in the previous studies.

The data of this study lend support to the hypothesis that strategies can mediate transfer to new problems. The data also suggest that young children can learn and transfer a complicated strategy. In this study second graders learned a nonreplacement strategy. They also transferred it to contexts substantially different from those used during training.

To develop a theory of instruction, data are needed about treatments which produce transfer. The teaching of problem-solving strategies to young children is both a way to obtain transfer and to develop a theory of instruction.

REFERENCES

- ANDERSON, R. C. Can first graders learn an advanced problem-solving skill? *Journal of Educational Psychology*, 1965, **56**, 283-294.
- ATKINSON, R. C., BOWER, G. H., & CROTHERS, E. J. *An introduction to mathematical learning theory*. New York: Wiley, 1966.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart & Co., 1960.
- MILLER, G. A. *Language and communication*. New York: McGraw-Hill, 1951.
- RESTLE, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329-342.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WITTRICK, M. C., & KEISLAR, E. R. Verbal cues in the transfer of concepts. *Journal of Educational Psychology*, 1965, **56**, 16-21.
- WITTRICK, M. C., KEISLAR, E. R., & STERN, C. Verbal cues in concept identification. *Journal of Educational Psychology*, 1964, **55**, 195-200.

(Received June 1, 1966)

SOME PSYCHOLOGICAL ASPECTS OF SUBJECT-MATTER STRUCTURE¹

PAUL E. JOHNSON²

University of Minnesota

A multiple-response verbal association test and a similarity judgments test were administered to 24 students of physics. The occurrence of words from the subject matter as responses on the association test was related to the frequency with which the words appeared in the written materials. Response distributions to words were similar for those words which were judged to represent similar concepts in the subject matter. Differences in achievement were related to the number of responses given per word on the association test as well as to the extent to which Ss' associations represented constraints in the subject matter.

Many of the concepts in a subject matter are defined by means of their relations with other concepts in that subject matter rather than by the presence or absence of specific environmental attributes (Carroll, 1964). This is especially true of physics, where much of the logical structure consists of interrelations among concepts which are specified formally by constraints within the written materials. Learning such a subject matter is in part a case of internalizing these relations, and one aspect of the psychological structure of physics consists of the learned interrelations among the words used to represent its concepts.

To the extent that a subject matter attempts to describe a certain por-

tion of human experience, which most do, there are also concepts present which are defined in terms of environmental data. Examples from this class of concepts in physics are operationally defined quantities such as *mass*, *distance*, and *time*.

The present paper is concerned with relationally defined concepts within the relatively stable subject matter domain of Newtonian Mechanics, and also with nonrelationally defined concepts within this domain insofar as they are elements of relations used to define other concepts. The purpose of the paper is to examine the associative relations that exist among the verbal labels for these kinds of concepts and then to consider the information that such relations provide about the structure of mechanics for individuals who are judged to have reached different levels of achievement in the subject matter.

In two previous papers (Johnson, 1964, 1965) it has been shown that the distributions of verbal associations to the words in a subject matter (the associative meanings of the words) reflect at least in part the constraints of the subject matter. However, words which represent concepts in mechanics may be associa-

¹This work was supported in part by grants in the University of Minnesota Center for Research in Human Learning, from the National Science Foundation (GS-541), the National Institute of Child Health and Human Development (HD-01136-01) and from the Graduate School of the University of Minnesota. Portions of this paper were read at the annual convention of the American Educational Research Association, February 1966.

²The author would like to express his appreciation to the chairman and his staff in the Science Department at the University of Minnesota Laboratory School for making this experiment possible.

tively related not only in the sense that one word elicits the other and vice versa, but also in the sense that as stimulus words they have responses in common. Words with overlapping response distributions have associative meanings which are similar in some degree and an index of this similarity among pairs of words is the overlap or intersection coefficient (Deese, 1962).

The structure of a subject matter is usually such that its concepts occur in more than one logical relation or constraint. Both associative meaning and associative overlap, however, are based upon procedures in which each subject (*S*) gives only a single response to each word on the association test. Once an *S* produces a response to a word, he is not permitted to give other responses to that word which may also be readily available and which would indicate learned associations involving the given word in addition to the relation suggested by the first response. Allowing each *S* to give only one response to each stimulus word also eliminates from consideration less available responses which may reflect learned relations among concepts that are utilized in more complex problem-solving tasks where the dominant associations are not appropriate.

A modification of associative meaning which has been adopted in the present investigation is one in which the associative meaning of a word is different for each *S* and is defined for each word as the ordered set of responses given by an *S* to that word in a fixed amount of time. The index of the degree of associative overlap between pairs of words is now termed a relatedness coefficient (Garskoff & Houston, 1963) and is a function of both the responses in common and their relative positions within their respective response hierarchies.

Associations learned among words as a result of their contiguous occurrence in the constraints of a subject matter may relate concepts which are, at least for the beginner, quite different things. And students learning a subject such as Newtonian Mechanics for the first time may make judgments about the conceptual similarity of words such as *mass* and *velocity* which are quite different from determinations made of their associative similarity by means of verbal association.

A second measure of psychological relation was therefore considered in the present investigation in order to compare the associative similarity of words with a judgment of the similarity of the concepts represented by the words.

METHOD

The *Ss* in this investigation were 24 seniors, 16 male and 8 female, from the University of Minnesota Laboratory School. The *Ss* were divided into two groups on the basis of their achievement in Newtonian Mechanics. The 12 *Ss* whose course grades were in the upper one-half of the distribution of grades were termed the high achievers and the 12 students whose grades were in the lower one-half were termed the low achievers. Similar proportions of each sex occurred in each achievement group.

A test booklet was prepared containing a page of general instructions, a continued word association test, and a similarity judgments test, in that order. The general instructions given to each *S* on the first page of the booklet were as follows:

The following pages contain test items designed to find out how people use words which represent concepts in physics. There are two tests, neither of which has anything to do with ability in physics, but which are intended to reveal facts about the usage of physics words not discoverable by more usual techniques. Treat each item on both tests as carefully as you can, even when the item calls for a judgment that seems almost impossible to make. Work rapidly, however, because we will learn most about how physics words are used if you give us the *first response which occurs to you*. You may open the booklet

and begin with the first page. Read the instructions carefully.

As a basis for selecting the stimulus words to be used in the present investigation, a frequency count was made of the words which represented concepts in Newtonian Mechanics in Ss' textbook (Dull, Metcalf, & Williams, 1960). This word count distinguished between the occurrence of a word representing a concept in running text, and the occurrence of the same word in all other situations (e.g., problems, questions, etc.). Fourteen words were selected from this word count so as to represent as much of the frequency range as possible, under the restriction that the list of words include six concepts whose definitions within the text consisted of simple physical equations (Johnson, 1965). The frequency range represented by the 14 words was from 1.26 occurrences per 100 words of running text for the word *FORCE* (one of the most frequently occurring concepts) to .02 occurrences per 100 words of text for the word *IMPULSE* (one of the least frequently occurring concepts). The rank-order correlation between the frequency of occurrence of these 14 words in running text and their frequency of occurrence in all other situations was .94, $p < .001$. The 14 words ranked according to their frequency of occurrence are presented later in Tables 3 and 4.

Each of the 14 words appeared 24 times on a page in the association test and each S's association test contained an individual random order of the 14 words. The instructions given to each S in the association test were as follows:

This is a test to see how many words you can think of and write down in a short time. You will be given a key word which represents a concept in physics and you are to write down as many other words which this key word brings to mind as you can. No one is expected to fill in all the spaces on a page, but write as many words as you can which the key word brings to mind. Be sure to think back to the key word after each word you write down because the test is to see how many other words the key word makes you think of. A good way to do this is to repeat the key word over and over to yourself as you write. You will have one minute on each page. I will tell you when to go on to the next page. Are there any questions?

Each of the 14 words on the association test was paired with every other word to create 91 pairs of words which served as stimuli on the similarity judgments test. Each pair of words on this test was followed

by a 7-point rating scale, the anchors of which were the words *similar* and *dissimilar*.

Two different random orders of the 91 pairs of words were constructed with the restrictions that (a) no pair of words be adjacent to the same pairs of words in the two random orders, and (b) no adjacent pair contain the same concepts. Word pairs were reversed within each random order so that each word was judged with every other word both when it appeared before and when it appeared after the word. The different orders of the pairs of words were separated within each random order so that no booklet contained a given pair of words more than once.

Ten pairs of words were printed on each page of the test booklet, except for one page which contained 11 pairs of words. Pages for both orders of the pairs of words were assembled in six different random orders, thus creating a total of 24 different forms of the judgments test. Six Ss at each level of achievement received each random order of words, two Ss at each level received each random order of the pages, and each S at each level received one of the two orders of the pairs of words. All randomization for the investigation was carried out by means of a table of random numbers.

The instructions given to each S on the similarity judgments test were as follows:

On the following pages you will find pairs of words which represent concepts in physics. Each pair of words is followed by a rating scale. These rating scales are designed to see how similar or dissimilar you feel the concepts represented by the words are to one another. If you feel the concepts represented by a particular pair of words are similar in some degree, then place an "X" in a blank near the "Similar" end of the rating scale. If, on the other hand, you feel that the concepts represented by the words are dissimilar in some degree, place an "X" in a blank near the "Dissimilar" end of the scale. The degree of similarity or dissimilarity is indicated by how far you place your "X" from either end of the scale.

The Ss were timed in the association test and allowed to proceed at their own speed on the judgments test. All Ss finished the booklet in one 45-minute class period.

RESULTS

The words on the association test were divided into two categories according to their frequency of occur-

rence in the subject matter; those words that occurred with a frequency greater than the median frequency for the 14 words of .11 occurrences per 100 words of running text and those words that occurred with a frequency less than this median frequency. A comparison was made between high and low achievers based upon the number of responses they produced to words in these two categories. The mean number of responses given by each type of *S* to words in each of the two categories is presented in Table 1.

An analysis of variance with repeated measures on the word count factor (Winer, 1962) was performed on the number of responses given by the high and low achievers to the words in these two categories. The high achievers gave significantly more responses per stimulus word than the low achievers ($F(1,22) = 6.79, p < .05$), and the high-frequency words elicited significantly more responses from *Ss* than the low-frequency words ($F(1,22) = 6.87, p < .05$). The interaction between achievement and word-count frequency was not significant.

One index of the psychological importance of words is the frequency with which they occur as responses on an association test. An index of the logical importance of the concepts in a subject matter is the relative frequency with which their verbal labels

occur in the materials used to present it. A determination was therefore made of the relation between the logical and psychological importance of words used to represent concepts in Newtonian Mechanics.

The 14 words used as stimuli on the association test were ranked according to their frequency of occurrence in the subject matter of mechanics. Each of the 14 words was also ranked on the basis of how often it was given as a response by the high and low achievers in the association test. The rank-order correlation between these two rankings of the 14 words was .87 ($p < .001$) for the high achievers and .52 ($p > .05$) for the low achievers.

Not only may words be psychologically important in the sense that they are given as responses to many different words, but also in the sense that when they occur as responses they do so near the top of *Ss'* response hierarchies. Each of the 14 words was therefore ranked according to its average position within the response hierarchies of the words on the association test. Separate rankings were made for high and low achievers and these rankings were each correlated with the ranking of the 14 words based upon their frequency of occurrence in the subject matter. This correlation was .52 ($p > .05$) for the high achievers and $-.05$ ($p > .05$) for the low achievers.

A number of the words on the association test represented concepts which are defined by simple equation relations in the subject matter of mechanics. These relations are $\text{FORCE} = \text{MASS} \times \text{ACCELERATION}$, $\text{ACCELERATION} = \text{VELOCITY}/\text{TIME}$, $\text{VELOCITY} = \text{DISTANCE}/\text{TIME}$, $\text{WORK} = \text{FORCE} \times \text{DISTANCE}$, $\text{POWER} = \text{WORK}/\text{TIME}$, and $\text{MOMENTUM} = \text{MASS} \times \text{VELOCITY}$. A comparison was made between high and

TABLE 1
MEAN NUMBER OF RESPONSES PER WORD
ON THE ASSOCIATION TEST

Word count frequency	Performance in the subject matter			
	High achievers		Low achievers	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
High	9.64	3.97	6.71	2.38
Low	8.86	3.84	6.24	2.34

low achievers based upon the number of their associations, termed constrained associations, to the nine words in the above six equations which were consistent with the relations specified by the equations. Thus, VELOCITY was a constrained associate of both ACCELERATION and DISTANCE whereas POWER was not a constrained associate of either of them.

Each *S*'s response hierarchy to each of the nine words in these six equations was divided in half and the number of constrained associations in each half was tabulated separately. The mean number of constrained associations in each portion of the response hierarchy for each type of *S* is presented in Table 2.

An analysis of variance with repeated measures on the response hierarchy factor was performed on the number of constrained associations given by the high and low achievers in the upper and lower half of their response hierarchies. The high achievers gave significantly more constrained associations to these words than the low achievers ($F(1,22) = 15.85, p < .001$), and there were significantly more associations of this kind in the upper than in the lower one-half of the *S*'s response hierarchies ($F(1,22) = 69.4, p < .001$). The interaction between achievement and hierarchy was also significant ($F(1,22) = 14.4, p < .001$). That is, there was a greater difference in the mean number of constrained associations in the two portions of the response hierarchy for the high achievers than there was for the low achievers.

In order to determine the associative similarity of words in the subject matter of Newtonian Mechanics, relatedness coefficients were calculated among all pairs of the words in the association test. The calculation of

TABLE 2
MEAN NUMBER OF CONSTRAINED
ASSOCIATIONS PER
STIMULUS WORD

Position within response hierarchy	Performance in the subject matter			
	High achievers		Low achievers	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Upper one-half	1.39	.92	.69	.73
Lower one-half	.43	.60	.32	.56

these coefficients was based upon the total distribution of responses given by each *S* to each of the 14 words rather than upon a subset of this distribution constructed from logical categories in the subject matter.

A tabulation was also made of each *S*'s similarity judgments to each pair of words. Median similarity judgments and median relatedness coefficients were computed separately for high and low achievers. The median relatedness coefficients for the high achievers are presented above the main diagonal in Table 3 and the median relatedness coefficients for the low achievers are presented below this diagonal.

The median similarity judgments were based upon a value of 7 for the similar end of the scale and a value of 1 for the dissimilar end of the scale. These values are presented above the main diagonal in Table 4 for the high achievers and below for the low achievers.

The rank-order correlation corrected for ties (Siegel, 1956) between a ranking of the 91 pairs of words according to median relatedness coefficients and a ranking of the word pairs according to median similarity judgments was .75 ($p < .001$) for the high achievers and .65 ($p < .001$) for the low achievers.

The 91 pairs of words were ranked

TABLE 3
MEDIAN RELATEDNESS COEFFICIENTS FOR
HIGH ACHIEVERS (ABOVE DIAGONAL) AND
LOW ACHIEVERS (BELOW DIAGONAL)

Stimulus words	Stimulus words													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. FORCE		25	41	24	36	32	09	11	16	34	30	20	34	21
2. VELOCITY	05		55	08	02	18	35	31	51	16	15	16	37	22
3. ACCELERATION	05	39		29	21	19	32	28	42	18	20	22	25	23
4. MASS	27	01	01		53	18	04	03	08	11	20	31	24	07
5. WEIGHT	24	03	0	47		17	02	03	04	11	08	18	24	07
6. WORK	38	06	56	07	06		26	29	13	55	40	17	24	17
7. DISTANCE	06	18	04	01	03	13		27	36	20	08	09	16	17
8. TIME	01	26	02	01	01	22	26		23	33	14	05	16	15
9. SPEED	03	61	47	01	15	30	20	10		11	14	24	21	
10. POWER	36	04	01	12	11	56	01	02	02		41	15	14	17
11. ENERGY	26	04	03	03	02	33	01	01	01	32		23	24	23
12. INERTIA	26	09	08	05	17	11	02	13	01	23	23		28	24
13. MOMENTUM	34	29	25	17	16	04	07	31	11	12	47	31		
14. IMPULSE	07	01	03	03	05	09	0	01	03	10	06	0	17	

Note.—Decimals omitted.

separately according to the magnitude of the median relatedness coefficients for the high and low achievers. This was done to determine the degree of agreement in the relative associative similarity of the word pairs for the two groups of Ss. The rank-order correlation corrected for ties between these two rankings of the pairs of words was .66 ($p < .001$).

The 91 pairs of words were also ranked separately according to the magnitude of the median similarity judgment for the high and low achievers. This was done to determine the degree of agreement in the relative judged similarity of the 91 pairs of words for Ss in the two groups. The rank-order correlation corrected for ties between these two rankings of the word pairs was .60 ($p < .001$).

Each of the words in Tables 3 and 4 was paired with every other word to create a total of 13 pairs for each of the 14 words. Rank-order correlations corrected for ties were then computed between median relatedness coefficients and median similarity judgments across the 13 pairs for each of the 14 words. This was done to determine the relative differences

among the 14 words in the extent to which associative and judged similarity were related to one another. These correlations were computed separately for high and low achievers and are presented in Table 5.

Each S's distribution of relatedness coefficients to the 91 pairs of words was divided into two categories; those coefficients which exceeded the median coefficient of the distribution for that S and those coefficients which did not exceed this median. In a similar manner, the distribution of similarity judgments for each S was divided into those judgments that exceeded the median judgment for the S and those that did not. A phi coefficient was then computed for each S based upon the two categories of relatedness coefficients and the two categories of similarity judgments. This was done to determine the degree of relation between the associative and judged similarity of words representing concepts in mechanics for each of the 24 Ss. The values of these 24 coefficients ranged from .007 to .457 with 13 significant at the .05 level ($r \geq .20$), 9 significant at the .01 level ($r \geq .27$),

TABLE 4
MEDIAN SIMILARITY JUDGMENTS FOR HIGH
ACHIEVERS (ABOVE DIAGONAL) AND LOW
ACHIEVERS (BELOW DIAGONAL)

Stimulus words	Stimulus words													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. FORCE		49	60	51	55	58	25	29	42	52	45	50	49	46
2. VELOCITY	30		55	35	25	29	48	58	62	28	45	45	55	40
3. ACCELERATION	48	57		55	57	40	40	58	56	45	45	45	35	35
4. MASS	40	38	23		55	57	14	14	22	40	45	47	55	38
5. WEIGHT	43	27	18	63		45	14	12	25	35	33	43	40	25
6. WORK	55	25	30	33	49		37	40	28	57	51	32	29	43
7. DISTANCE	23	32	30	15	18	35		35	46	35	25	30	32	40
8. TIME	20	40	50	17	18	25	25		45	52	35	30	35	40
9. SPEED	30	64	59	35	25	33	34	52		35	38	39	45	30
10. POWER	58	30	48	35	37	63	30	32	45		45	39	33	35
11. ENERGY	52	32	42	25	30	57	25	21	40	57		45	45	44
12. INERTIA	43	30	30	30	43	40	21	25	45	38	42		60	31
13. MOMENTUM	50	49	55	53	52	35	23	23	53	40	44	50		35
14. IMPULSE	38	45	25	35	25	25	24	28	38	25	45	25	40	

Note.—Decimals omitted.

TABLE 5

RANK ORDER CORRELATIONS BETWEEN RELATEDNESS COEFFICIENTS AND SIMILARITY JUDGMENTS FOR EACH WORD WITH ALL OTHER WORDS ON THE ASSOCIATION TEST

Stimulus word	Performance in the subject matter	
	High achievers	Low achievers
FORCE	.81***	.83***
VELOCITY	.82***	.50
ACCELERATION	.41	.54
MASS	.82***	.69**
WEIGHT	.90***	.85***
WORK	.44	.38
DISTANCE	.94***	.46
TIME	.80***	.56*
SPEED	.91***	.55*
POWER	.67**	.55*
ENERGY	.73**	.88***
INERTIA	.68**	.66*
MOMENTUM	.59*	.72**
IMPULSE	.03	.27

* $p < .05$.

** $p < .01$.

*** $p < .001$.

and 6 significant at .001 level ($r \geq .34$).

Finally Ss were divided into two groups according to the degree of relation between their relatedness coefficients and similarity judgments for the 91 word pairs. These groups consisted of those Ss whose phi coefficient exceeded the median coefficient for the combined groups and those Ss whose phi coefficient did not exceed this median. The division of Ss was compared with the two-category division of Ss made on the basis of achievement in the subject matter. A chi-square test indicated no significant relation between achievement and degree of relation between associative and judged similarity across word pairs representing concepts in the subject ($\chi^2 = .68, p > .05$).

DISCUSSION

Words which represent concepts in physics are related to one another by constraints inside as well as outside the domain of the subject matter. One form of relation which may exist among words due to their contiguous occurrence in such constraints is that of association.

An examination of the associative relations among words which represent concepts in Newtonian Mechanics indicates that those words which occurred frequently in the written materials were more meaningful for both high and low achievers than words which occurred infrequently. In addition, both frequently and infrequently occurring words were more meaningful for high achievers than they were for low achievers.

The occurrence of words in mechanics as responses on an association test was also related, for high achievers, to the relative frequency with which these words occurred in the subject matter. The relative order in which words in mechanics occurred within the response hierarchies of other words in the subject matter was not, however, related to the frequency of occurrence of the response words in the written materials of mechanics for either high or low achievers.

One form of subject-matter constraint which serves to define concepts in mechanics is the physical equation. This form of constraint was represented in association more frequently for high achievers than for low achievers. To the extent that relations appeared in association which were consistent with such constraints, they occurred early in Ss' response hierarchies, especially for the high achievers.

Judgments of the similarity of the concepts in Table 4 were related to

the associative similarity of the words representing these concepts as indexed by the coefficients in Table 3, and the degree of this relation was high for both high and low achievers. Furthermore, the relative associative similarity among the pairs of words in Table 3 was similar for the two groups of Ss, as was the relative judged similarity of the concepts in Table 4.

The data presented in Table 5 indicate that the degree of relation between associative and judged similarity across the word pairs involving a single concept depended upon both the concept and the group of Ss considered. The highest degree of relation for the high achievers occurred in those pairs in which one of the concepts was DISTANCE. For the low achievers, the highest degree of relation occurred for the pairs involving the concept ENERGY. The three concepts for which there was least agreement among the high achievers as to the degree of relation between associative and judged similarity were WORK, ACCELERATION, and IMPULSE, while the three concepts for which there was least agreement among the low achievers were WORK, DISTANCE, and IMPULSE. The concepts for which there was the largest difference between the two groups in favor of the high achievers were VELOCITY, SPEED, DISTANCE, TIME, and MASS.

Students begin their formal study of physics with learned associations among the words used to represent many of its concepts due to the verbal environments in which these words have occurred outside the domain of the subject matter. Within the subject matter these words are related to one another and to other words by formal constraints which take the form of equations and definitions. The learning of these constraints pro-

vides additional associations which, together with the previously learned relations, form the basis for the associative similarity among pairs of words.

Judgments of conceptual similarity are also influenced by constraints both inside and outside the domain of the subject matter and associative and conceptual similarity may covary due to a mastery of the verbal relations which appear in these constraints. These judgments may, on the other hand, be unrelated to associative similarity either because the constraints are unavailable to Ss at the time of testing or because the associative relations representing the constraints are selected from one domain while judgments of conceptual similarity are based upon constraints in another.

Although in the present investigation the general degree of relation between associative and judged similarity was high for both high and low achievers, there was considerable variation between the two groups of Ss for individual concepts in the subject matter. The largest differences, which were in the direction of increasing competence, were for the operationally defined concepts of MASS, DISTANCE, and TIME, either as they occurred separately with other concepts or as they occurred in combination with one another in the constraint $VELOCITY = DISTANCE/TIME$.

Among the concepts in Newtonian Mechanics, operationally defined concepts have a particular status in that they alone serve to represent measurable properties of objects in a student's environment. One result of mastering mechanics is that such concepts can be dealt with in terms of the subject-matter constraints in which they occur rather than in terms of their more general environmental meanings

determined by past experience. Perhaps knowledge of the structure of a subject such as Newtonian Mechanics would proceed more rapidly than it does if more of the words which represent its concepts were not related to outside patterns of usage but were instead learned simply as they are related to one another by means of the formal constraints in the subject matter.

REFERENCES

- CARROLL, J. B. Words, meanings and concepts. *Harvard Educational Review*, 1964, 34, 178-202.
- DEESE, J. On the structure of associative meaning. *Psychological Review*, 1962, 69, 161-175.
- DULL, C., METCALFE, H. C., & WILLIAMS, J. E. *Modern physics*. New York: Henry Holt, 1960.
- GARSKOFF, B. E., & HOUSTON, J. P. Measurement of verbal relatedness: An idiographic approach. *Psychological Review*, 1963, 70, 277-288.
- JOHNSON, P. E. Associative meaning of concepts in physics. *Journal of Educational Psychology*, 1964, 55, 84-88.
- JOHNSON, P. E. Word relatedness and problem solving in high-school physics. *Journal of Educational Psychology*, 1965, 56, 217-224.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw Hill, 1956.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received April 14, 1966)

FREE-OPERANT PREFERENCE FOR ONE OF TWO STORIES:

A METHODOLOGICAL NOTE

THOMAS C. LOVITT

University of Washington

2 recorded stories were simultaneously presented to 7 12-yr.-old children. By programming these stories by conjugate reinforcement procedures S was able to select continuously the story of his choice by either pressing or not pressing a hand switch. These hand-switch responses were graphically recorded, thus enabling E to determine not only the individually preferred story, but the quantitative amount this option represented. The advantages of the conjugate technique in obtaining quantifiable preference data and future educational involvement of the conjugate procedure are discussed.

Previous investigations have revealed that by conjugately or continuously offering a stimulus, either visual or aural, a subject's (S's) interest at any given time could be measured. Lindsley (1962), in analyzing the visual narrative effects of specified television programs, demonstrated that, for his experimental Ss, rerun movies maintained a more rapid and consistent rate of response than did programs especially designed for television viewing. In a study involving the comparative effects of five classes of verbal narration, Lovitt (1965) reported that although a conventional story generated higher and more consistent responding to acquire the narration, other and more abstruse verbalizations were more desirable than silence.

Based on the evidence of these two investigations, which compared visual or aural narration to no stimulation, an experiment was designed to contrast the reinforcing effects of two simultaneously offered stories. In addition to obtaining each individual's operant preference for either story, information would also be gathered relevant to his verbal choice. This latter concern was prompted by Morgan and Lindsley (1964), who compared individual operant and verbal preferences and reported that these

dimensions of preference were not always synonymous. Their study, which involved the simultaneous offering of monophonic and stereophonic music, demonstrated that while actively consuming the stimuli some Ss preferred stereo music, whereas others demonstrated no preference. However, after the session, when the stimuli were no longer available, all Ss verbalized a choice for stereophonic music.

METHOD

Subjects

The participants in this study were five boys and two girls, all 12 years of age. All of the Ss were enrolled in either seventh- or eighth-grade classes in Northeast Johnson County, Kansas.

Procedure

All of the Ss participated in one experimental session lasting from 30 minutes to 1 hour. During this session, each child sat in a partially sound-attenuated room furnished with only a chair, a set of headphones, and a hand microswitch. He was told only that he would be listening to stories and that he could press the switch if he so desired. Nothing was said about the reinforcement contingencies—whether it was or was not necessary to press the switch, or if so, at what rate.

The microswitch, when pressed, produced a brief electric charge that affected the narrative intensity. Even if the switch was

pressed and held down, only this brief intensity change resulted. The *response definer* converted each of these microswitch presses into electrical pulses that in turn operated the *conjugate reinforcer*.

Two stories, which had been previously recorded by the investigator, were programmed by the conjugate-servo. They were begun at the same time and ran continuously throughout the experimental session.

The stories available to each *S* were *The Incredible Journey* by Sheila Burnford (1961) and *White Falcon* by Elliott Arnold (1955). Both stories have been used frequently in classrooms and homes for 12-year-olds and have received such commendation as reinforcing literature for the preteen as being presented the William Allen White book award.

In addition to their suitability for preadolescents, these stories met another criterion to encourage the factor of selection—they offered both constant narrative action and a certain amount of content redundancy. The former was a necessary ingredient, in that to present simultaneously two stories in a measure of preference, it is imperative that the narrative content of both stories be fairly equally reinforcing throughout. The matter of redundancy was an equally important aspect in that during some experimental contingencies the preferred story would be unobtainable, thus leaving a gap of a few minutes in its presentation.

In order to attain the desired story at the preset maximal volume (which was set at a comfortable level) during experimental accelerating sessions, the *S* had to respond at a rate of 45 responses per minute. If he responded at a rate lower than this, the story gradually faded, and he was presented with portions of both stories at minimal volumes. If *S* did not respond at all, he was granted the alternate story, which was programmed by decelerating contingencies.

The basic sequential order of each session consisted of the following segments:

J	W	W	J	J	W
<	<	<	<	<	<
W	J	W	J	W	J
— 1	— 2	— 3	— 4	— 5	— 6 —

The J was used as the code letter for *Incredible Journey*; the W to denote *White Falcon*. The top letter indicates accelerating conditions, under which *S* must emit a required response rate in order to obtain the

stimulus at maximum volume. The bottom letter indicates decelerating conditions, whereby *S* need not respond to acquire that stimulus. In Segments 3 and 4 (control conditions) *S* was granted either *White Falcon* or *Incredible Journey*, regardless of responding, not responding, or rate of response. He was unable to avoid the stimulus during these conditions.

Each listener's operant response rate was recorded on a cumulative response recorder. This equipment is comprised of a roll of paper moving continuously at 30 centimeters (about 12 inches) per hour and a recording pen that records each handswitch response by an upward movement. After 500 of these recorded responses, the pen automatically resets and the moving process is repeated. The slope of the line on these cumulative records indicates the rate of pressing the handswitch, a steep slope indicating rapid responding and a smooth slope indicating continuous responding. These cumulative records enable the experimenter to measure directly the listener's moment-to-moment interest, or his desire to listen to either story.

At the conclusion of the session each *S* was asked to verbalize his preference. This stated preference was then compared to his operant choice.

RESULTS

The upper segment of Figure 1 is a cumulative record of an *S* who evidenced an operant preference for J over W. This *S*, referred to as *S* 1, was the only child in this experiment who expressed a preference for, and operated to attain, J.

Her operant preference for J was unequivocal in that she responded to acquire this narrative form at a high and consistent rate during Segments 1 and 5 when J was programmed as an accelerating consequence. She further evidenced her J option by not responding when J was programmed as a decelerating consequence during Segments 2 and 6. Final confirmation of a J choice was demonstrated during Segment 4, when she accepted J without responding, and during Segment 3, when she performed in a searching or probing manner in a vain attempt to

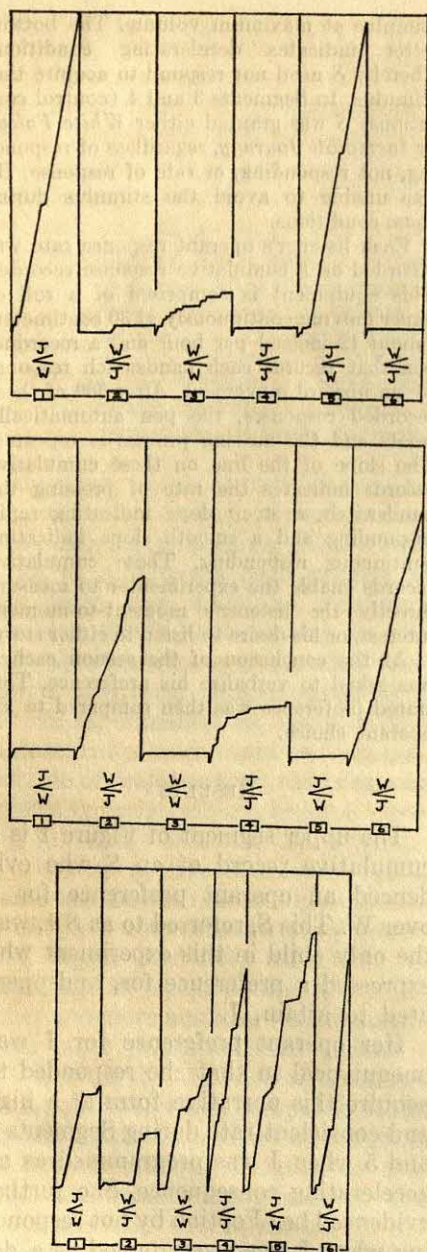


FIG. 1. The upper segment demonstrates *S* 1's operant preference for J. (The center graph depicts *S* 2's operant preference for W. The lower segment demonstrates *S* 5's slight operant preference for W.)

be granted the more favored, but unobtainable, J.

S 2's operant preference for W over J was representative of the performance of four other *Ss*. The center portion of Figure 1 is a cumulative record of his response behavior.

Although *S* 2's response rate to acquire his favored stimulus was lower than that of *S* 1 to acquire her favored narration, it can be noted by comparing their records that their response rate patterns are complete opposites. During Segments 1 and 5, when W was programmed as a decelerating consequence, *S* 2 refrained from responding, accepting W without effort. However, when W was programmed as an accelerating consequence, during Segments 2 and 6, he met the contingency requirement by responding beyond the minimal setting of 45 per minute. During Segment 3, a control segment, where W was programmed as both a decelerating and accelerating consequence, *S* 2 accepted the "free" offering and did not respond. Final confirmation of a W choice was indicated when he responded in a searching, yet futile, manner during Segment 4 to obtain his favored W. During Segment 4, *S* 2's irregular and intermittent response pattern was similar to that of several other *Ss* in this research during the nonpreferred control segment, since at this time the favored narration was unobtainable.

The lower portion of Figure 1 is the cumulative record of *S* 5, the only *S* who did not perform in such a way that a definitive operant selection could be ascertained. Although *S* 5 verbally expressed a preference for W over J, this fact was not totally substantiated by his ongoing operant choice. Observation of his cumulative record reveals that *S* 5 did some responding during all experimental segments, including the control portions (Segments 3 and 4), where responding of any type was completely unproduc-

tive. If *S* 5's response rates during the final two segments (5 and 6) are compared, some slight preference for *W* is noted in that he worked at 78 per minute to attain *J* and 130 to acquire *W*. However, if the response rates during Segments 1 and 2 are compared, an alternate interpretation may be made, in that he responded more for *J* during Segment 1 than for *W* during Segment 2.

As a group the data revealed that five *Ss* operated to acquire *W*, one to obtain *J*, and one to be granted portions of each.

When verbal preferences were compared to operant choices, five consumers had synonymous reactions; whereas in two instances the operant and verbal choices were not in agreement. *S* 5 had expressed a preference for *W* and operated to acquire portions of *J* as well as *W*; while *S* 6 responded consistently to obtain *W*, but verbalized no particular preference for either story.

In view of the fact that in this study all but two of the *Ss*' verbal and operant preferences coincided, obtaining verbal options alone would have been a reasonably valid indicator of preference. One might further speculate that for this sample of children their ongoing, or operant, preference was virtually the same as their verbalized choice made after the narration had

been consumed. However, this stated or verbalized choice could be useful only if an either-or dimension of data is sufficient. On the other hand, the continuous operant preference, since it has a numerical equivalent, enables the experimenter to make quantifiable statements pertinent to the relative interest of one story as compared to another.

Since the effectiveness of the conjugate methodology, insofar as continuously measuring individual aesthetic preference is concerned, has been demonstrated, implications for further application of this technique seem rather obvious. Not only might this tactic prove useful in an educational setting, but future research may validate its effectiveness in broader areas of communication and experimental esthetics.

REFERENCES

- ARNOLD, E. *White falcon*. New York: Knopf, 1955.
- BURNFORD, S. *The incredible journey*. Boston: Little, Brown, 1961.
- LINDSLEY, O. R. A behavioral measure of television viewing. *Journal of Advertising Research*. 1962, 2, 2-13.
- LOVITT, T. C. Operant preference for narration over silence. Unpublished manuscript, University of Washington, 1965.
- MORGAN, B. & LINDSLEY, O. R. Operant preference for stereophonic over monophonic music. Unpublished manuscript, University of Kansas, 1964.

(Received April 26, 1966)

SOCIAL SCHEMA OF NORMAL AND DISTURBED SCHOOL CHILDREN¹

RHODA LEE FISHER²

Syracuse Scholastic Rehabilitation Program, Syracuse, New York

The Kuethe social schema technique significantly differentiated between emotionally disturbed and normal boys of elementary school age. The difference was such that the disturbed boys placed greater distance between figures in the social schemas. The degree of distance the disturbed boys place between the figures was positively correlated with the amount of hostility shown by their mothers. Mother's hostility was measured with the Buss-Durkee Hostility Scale. It would appear that children with hostile mothers conceive of social relations as more distant than children whose mothers are less hostile.

The present studies concerned themselves with a group of children whose deportment was so disruptive that teachers could not exercise sufficient control over them to permit them to remain within the normal classroom. The inability to behave in an acceptable socially approved fashion seems to be a critical problem for them. A measure of social distance has been developed by Kuethe (1962) which schematically represents the individual's concept of social interaction. With this technique he found that

individuals who employ a specific social schema in organizing behavior in one situation will employ the same schema in quite different situations where the one common denominator is social stimuli of the same content even though the physical form of the stimuli may be completely changed [Kuethe, 1964, p. 24].

Within the context of the application

¹ This study was partially supported by National Institutes of Health Grant MH-01475.

² Acknowledgement and thanks are expressed to the school principal, Frank Liss, and to teachers W. Liston, T. Clift, and E. Feeley, all of the Syracuse public schools, whose cooperation facilitated the collection of this data.

of the social schema technique, the following objectives were set:

1. To compare the social schema of normal children with those characterizing children with school behavior problems. It was anticipated that those showing disruptive behavior in the classroom would portray interaction schema differently than would normally behaving children. More specifically it was expected that the disturbed children would experience themselves as having fewer meaningful ties to others—as more alienated and without meaningful roles. That is, they feel separated and distant from others.

2. To relate distances set by children between human forms and the degree of hostility characteristics of their mothers. It was hypothesized that the more angry and aggressive the mother the more distantly would her child structure social relations. That is, one would expect that the angry mother would be perceived as repelling and pushing others away. She would exemplify a style of social relationship based on distance and unfriendliness.

3. To compare a group method of administration of social distance schema with an individual method of administration.

STUDY I

Method

Subjects. One group of subjects (*Ss*) consisted of 32 white male elementary school children in special classes designed for those unable to adjust to the regular school demands. These children were too difficult for the regular classroom teacher to control. The mean age of the group was 9.5 (range 6-13). A second group of *Ss* was comprised of 45 white male elementary school children from the regular fifth- and sixth-grade classrooms; and a third group included 44 white female elementary school children from the regular classroom. The mean age for both normal groups was 11.2 (range 10.0-13.0).

Procedure

An adaptation of a technique developed by Kuethe (1962) for the measurement of social schemata was used. Instead of Kuethe's usual felt figures with directions to place them on a large felt background mounted on a wall, the instructions were to glue the figures on an open field (in a booklet of paper). This shift in procedure enabled the series to be administered on a group basis. Specifically, each sample of *Ss* placed eight groups of figures on a series of fields according to Kuethe's (1962) "free response instructions." The figures for the group series consisted of yellow cutouts on gum-backed paper varying in height from 2½ inches (the baby figure) to 5½ inches (the man figure). Each *S* was provided with a booklet consisting of eight pages of 6-inch by 22-inch paper. An envelope containing figures to be placed on each page was distributed to each *S*. Not until everyone had completed his arrangement of the figures on a given page were instructions given to turn to the next page; and the next envelope was distributed. A description of the figures contained in each envelope follows:

Envelope 1: two boys.

Envelope 2: a girl, a woman.

Envelope 3: a woman, a man.

Envelope 4: two boys, one girl.

Envelope 5: a man, a girl.

Envelope 6: a man, a boy.

Envelope 7: a boy, a girl.

Envelope 8: a boy, a girl, a baby.

Scoring consisted of measuring the distance between each two figures. The score derived was the average distance for the entire eight situations.

Results and Discussion

The average distance between the figures arranged in schemas was significantly different for the disturbed and the normal samples. Normal boys arranged the human figures more closely together than did the disturbed boys. Mean total average distance for the normal boys was 1.6 inches; mean total average distance for the group of disturbed boys was 2.3 inches ($t = 2.9, p < .01$).

A comparison between the normal boys and the normal girls failed to reveal a real difference. The total average distance of the figures as grouped by the girls was 1.6 inches.

Since no sex differences were apparent between the normal boys and the normal girls it seemed logical also to compare the normal girls with the disturbed boys. When this was done, a significant difference was once again found ($t = 2.3, p < .05$).

Another analysis in which age was controlled indicated that age exerted no influence on the distance between figures as represented in the social schemata. Children below the age of 10.0 were eliminated from the disturbed group. Children above the age of 11.0 were eliminated from the normal group of boys. The mean age of both groups for this analysis was 10.4 years. Mean total distance between figure placement for the normal boys ($N = 12$) was 1.4 inches. The closeness of these values to those for the total samples indicated that age did not influence the perception of social distance in the present experimental context.

As predicted, normal children placed figures with human characteristics closer together than did children with serious school problems. In terms of Kuethe's work this suggests that the normal children feel closer

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE
NORMAL AND DISTURBED CHILDREN ON
THE GROUP ADMINISTERED FORM OF
THE SOCIAL SCHEMATA TECHNIQUE

Subjects	Total average distance		Average distance between child figures	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Disturbed boys	2.3	1.2	2.3	.7
Normal boys	1.6	1.0	1.7	1.1
Normal girls	1.6	.9	1.9	1.4

and more related to others than do the disturbed children.

In a paper recently published Weinstein (1965) also reports that the Kuethé schema placements discriminate disturbed and normal boys. The figures and felt size used by Weinstein were not directly comparable to those employed in the present study. In her study a group of 20 disturbed boys admitted to a special residential school for emotionally disturbed children was evaluated. A group of 20 normal control children from a nearby public school matched in age and IQ was used as a comparison group. Results indicated that disturbed boys placed figures that represented mother-child relationships at a relatively greater distance from each other than did normal boys. A second finding indicated that disturbed boys, as compared to normal boys, visualized human relationships as involving more intervening distance than is true of relationships between abstract geometric forms. Overall the Weinstein data and those from the present project, while not specifically supportive of each other, certainly indicate in common a trend for disturbed boys to conceptualize human relationships as distant.

STUDY II

Method

For this study only those boys in the special disturbed classes were used. It was the purpose of this project to investigate relationships which might exist between the hostility attributes characterizing the mother and the closeness-distance positions assigned to human figures on the Kuethé schemas by the children. Since aggression is such a prominent dimension in the behavior of the children it seems an important dimension to investigate in the mothers as well. To this end, five subscales of the Buss-Durkee Hostility Scale (Buss, 1961, pp. 171-172) were administered to each mother. These five scales measured Assaultiveness, Irritability, Negativism, Verbal aggressiveness, and Guilt. The *S* was asked to respond with a true or false answer to 48 different statements regarding the expression of anger. Examples of the Assaultive aggressive subscale are "Once in a while I cannot control my urge to harm others," "People who continually pester you are asking for a punch in the nose," "I have known people who pushed me so far that we came to blows." Examples of the Irritability scale are "If someone doesn't treat me right, I don't let it annoy me," "Sometimes people bother me just by being around," "I can't help being a little rude to people I don't like." The Negativism scale includes statements such as "Unless somebody asks me in a nice way, I don't do what they want," "When I disapprove of my friends' behavior, I let them know it," "I can't help getting into arguments when people disagree with me" are examples of the Verbal aggression subscale. The Guilt subscale includes statements such as "I sometimes have bad thoughts which make me feel ashamed of myself," "I am concerned about being forgiven for my sins."

An individual form of the social schemata was used in this experiment with the disturbed group of boys. The individually administered form consisted of seven groups of figures cut from yellow felt. The figures were derived from Kuethé's (1962, pp. 32-33) original patterns and *Ss* were instructed to arrange each group of figures on a black felt field 36 inches by 36 inches that was stretched on a wall of the experimental room. This series consisted of the following groups:

1. Two children and a square.
2. Two children and a woman.
3. Two children and a man.

4. A man, a woman, a square, and a large rectangle.
5. A man, a woman, a child.
6. Six children.
7. A child, a book, a circle

Results and Discussion

Four scores were obtained from the individually administered technique: the total average distance between all of the figures, the average distance between all of the adult figures, the average distance between all of the child figures, the average distance between adult and child figures.

Two measures from the Buss-Durkee Hostility scale were positively related to the distance placement between figures in the social schemata. The Buss-Durkee Assaultiveness scores of the mothers were positively correlated with the childrens' indicated distances between adult Kuethe figures ($r = .57, p < .01, N = 25$). The mothers' Irritability scores were positively correlated with the distances between adult Kuethe figures ($r = .53, p < .01, N = 25$). The total average social schemata distances

TABLE 2

CORRELATION OF MOTHER'S BUSS-DURKEE HOSTILITY SCALE SCORES WITH VARIOUS MEASURES OF KUETHE'S SOCIAL SCHEMATA

Buss-Durkee	Individually tested			Group tested
	Average adult distance ($M = 2.9$)	Average child distance ($M = 3.1$)	Total average distance ($M = 3.1$)	Total average distance ($M = 2.3$)
Direct aggression ($M = 3.6$)	.57***			
Irritability ($M = 5.2$)	.53***	.37*	.55***	
Negativism ($M = 1.8$)				-.68***

* $p = .05$.

*** $p = .01$.

TABLE 3

CORRELATIONS BETWEEN THE GROUP ADMINISTERED AND INDIVIDUALLY ADMINISTERED FORMS OF THE SOCIAL SCHEMATA TECHNIQUE

Individually administered form	Group administered form	
	Total average distance ($M = 2.3$)	Average distance between child figures ($M = 2.3$)
Average distance between adult figures ($M = 2.9$)	.43**	.37*
Average distance between adult and child figures ($M = 2.9$)	.09	.08
Average distance between child figures ($M = 3.1$)	.05	.03
Total average distance ($M = 3.1$)	.18	.12

* $p = .05$.

** $p = .02$.

were also positively correlated with the mothers' Irritability score ($r = .53, p < .01, N = 26$). Mothers' Irritability scores were also positively correlated with the distances between the child figures ($r = .37, p = .05, N = 26$).

Generally then, children who place human figures at a relatively large distance from each other have mothers who are depicted as angry and hostile.

STUDY III

Method

A comparison of the two forms (individual and group) of administration of the social schemata technique was the purpose of the third study. The children in the special classes for the disturbed were administered both versions of the Kuethe procedure. Though the situations for each of the series are not completely parallel, they seemed sufficiently similar to make such an analysis worthwhile.

Results and Discussion

Distance between adult figures for the individually-administered schemata was significantly related to the equivalent distance between figures for the group-administered schemata ($r = .43, p = .02, N = 30$).

Distance between adult figures for the individually-administered schemata was significantly related to distance between child figures for the group-administered schema ($r = .37, p = .05, N = 30$).

There is evidence indicating that a similar dimension is tapped by both versions of the Kuethe technique.

CONCLUSIONS

Children with problems in adjustment to the classroom framework place human figures at a significantly greater distance apart than do children who are able to adjust successfully to the classroom. In terms of Kuethe's findings (1962), this indicates that the disturbed children feel relatively distant or estranged from others. Whether this sense of distance is produced by the negative conse-

quences of their behavior or actually plays a previous role in contributing to their poor adjustment remains to be determined. There is the possibility that the degree of distance felt by a child between himself and others arises from socialization experiences. This is suggested by the fact that a significant positive correlation was found between Kuethe scores and measures of mother hostility in the disturbed group. Structuring one's relations in a "distant" fashion may be a consequence of dealing with a mother who angrily pushes him away or perhaps of needing to remain far enough away from mother to avoid her aggression.

REFERENCES

- BUSS, A. H. *The psychology of aggression*. New York: Wiley, 1961.
- KUETHE, J. Social schemas. *Journal of Abnormal and Social Psychology*, 1962, **64**, 31-38.
- KUETHE, J. Pervasive influence of social schemata. *Journal of Abnormal and Social Psychology*, 1964, **68**, 248-254.
- WEINSTEIN, L. Social schemata of emotionally disturbed boys. *Journal of Abnormal Psychology*, 1965, **70**, 457-461.

(Received May 5, 1966)

IDENTIFICATION PATTERNS, MOTIVATION, AND SCHOOL ACHIEVEMENT OF BRIGHT JUNIOR HIGH SCHOOL BOYS¹

THOMAS A. RINGNESS

University of Wisconsin

Interview and card sort data were obtained on 261 high-, average-, and low-achieving bright 8th-grade boys in an attempt to confirm or refute earlier findings concerning identification patterns, motivation, and values. Confirmation was generally found, the main exception being that in the present study most Ss identified with fathers whereas in the previous study high achievers were most likely to do so. Socio-economic status bias may have influenced earlier data. Low achievers were found more motivated than others to affiliate with peers; high achievers were more motivated academically. Low achievers were more nonconforming, whereas high achievers were more independent. School was seen by most Ss as demanding conformity, and Ss accepted this role model. Scholarship was shown to have little relationship to peer popularity, and the perceived norm for school achievement was that of mediocrity.

Identification may be defined as the process of affiliation with one or more other persons, groups, or institutions, which tend to become models. Attitudes, values, and other behavior are imitated, and may be internalized by the imitator.

Identification is closely related to motivation, since one tends to identify with those who provide him suitable reinforcers, and, on the other hand, reinforcers become so partly because they are dispensed by those with whom one identifies.

An extended discussion of identification and imitation is provided by Bandura and Walters (1963). germane to the present study is the generally accepted belief that, in early childhood, both boys and girls tend to identify most closely with the mother. After early childhood, however, the small boy tends to identify with the father and to accept him as

a role model; this transition usually takes place by the time the boy reaches school age.

In developing his system of values, attitudes, and motives, the child is exposed to other adult models (such as teachers), and to models in the peer group. He rarely reproduces all of the attributes of any given model, but selects certain ones from each. Accordingly, if identifying figures have many common attributes, there will be a corresponding reinforcing effect and the child's behavior should strongly reflect the consensual elements. If, however, identifying figures portray diverse behaviors, the child's own attributes may vary with the influence of the identifying figures.

In relation to motives to achieve in school, a number of possibilities exist. The boy may or may not identify closely with his father, or with both parents. The parents may or may not value school achievement and reinforce achievement striving. The boy may or may not identify with his teachers. The peer group with which he identifies may or may not value school achievement. Depending, there-

¹The research reported herein was supported by the Cooperative Research Program of the United States Office of Education, United States Department of Health, Education, and Welfare (Project S-035, Final Report 1965).

fore, on the identification patterns of the boy, the values of his models, those he has internalized, and his perception of school, motivation to achieve will vary. It is expected that these variables will be related to differences in academic achievement of children whose abilities and cultural backgrounds are such as to permit above-average school achievement.

In the present study no formal hypotheses were proposed, but questions were raised concerning the extent of the subjects' (Ss') identification with fathers, teachers, and peers, and with acceptance of school values, as these variables related to school achievement.

The research reported here is part of a larger study (Ringness, 1965c). It partially replicates an earlier study (Ringness, 1963, 1965a, 1965b) and is an attempt to confirm or refute previous findings. Certain differences in sampling and instrumentation were introduced so that the two studies are not entirely comparable.

The earlier research employed matched pairs of ninth-grade boys, half of whom attained a grade-point-average (GPA) of 3.00 or above (based on a system where A = 4.00) and half of whom attained a GPA of 2.00 or below. The remarks of Thorndike (1963), among others, suggest that failure to include the middle range of achievement might neglect possible curvilinear relationships among data; the present study attempts to compensate for that omission.

Further, the earlier research employed a sample of Ss from an upper-middle socioeconomic status (SES) population. Since parents and youth from this SES may be more oriented toward school achievement than those from other SES classes, there may have been a bias in the earlier data;

the present study embodies Ss from all SES classes.

Girls were not studied in either research, primarily because boys show poor school achievement approximately four times as frequently as girls, hence seem to constitute a more pressing problem. In addition, various studies (e.g., McGuire, Hindsman, King, & Jennings, 1961) have shown sex-specific factors related to school achievement.

RELATED LITERATURE

This study does not assume generalized motives to achieve (e.g., McClelland's, 1953, achievement motive). Rather, it is considered that behavior related to school achievement is related to values possessed and reinforcements offered by identifying figures. These may vary among themselves, with resultant influences on motives to achieve in school.

A number of studies have been concerned with the relationships of academic achievement to parent-child relationships. Taylor (1964) surveyed studies of personality traits related to discrepant school achievement. Employing the terms "overachiever" and "underachiever," he found general support for the belief that the overachiever accepts authority and has a good relationship with his parents. The parents tend to be supportive of their children's academic efforts and the children try to please them by doing well academically. Morrow and Wilson (1961) and the Portland Public Schools (1959) found that families of high achievers were more supportive, less authoritarian, and more permissive than families of low achievers; presumably such relationships encourage children's independence and striving behaviors.

However, the dynamics are not entirely clear, for Drews and Teahan

(1957) found that domineering mothers tend to foster high achievement in their children, and Taylor notes that some studies suggest that overachievers may be compensating for a lack of love, warmth, or understanding at home. In the latter event, identification with the home might be lacking, but identification with the school might be present.

In contrast to high or overachievers, low or underachievers tend to react against the home and school, and provide more problem behavior for authorities (Frankel, 1960; Gowan, 1955; Ringness, 1963). Taylor found underachievers to be usually regarded as hostile and aggressive toward authorities, and conflict with parents seemed to be carried over to authority figures outside the home. Bandura and Walters' (1959) work on adolescent aggression tends to confirm this belief, and supports the notion that for boys, the father-son relationship may be the most important determiner. The weight of evidence suggests that identification with the home, especially with the father, may be closely related to the school achievement of boys.

Motive to affiliate with peers has been studied extensively. Ringness (1963) found that low achievers were more motivated to affiliate with peers than were high achievers, in contrast to the family orientation and engagement in family activities characteristic of high achievers. Taylor's review found that overachievers were more independent, leaderlike, responsible, and dependable than underachievers; the latter tended to identify with a peer group, from which they gained support. Peer group affiliation was seen as a means of gaining some emotional security and as a way of reducing anxiety for underachievers.

McGuire et al. (1961) tend to con-

firm the notion that low achievers are more dependent on peer group support, but there is an implication that low achievers are not well accepted by the peer group at large. In a factor analytic and multiple regression study of predictors of talent, two independent peer-related factors were isolated. "Peer Stimulus Value" (PSV) was defined as a positive response to pressures imposed by age-mates. Children high in this factor were regarded as models by their peers; they were active, accepted, self-confident, and effective. A person high in the second peer-related factor, "Age-Mate Avoidance," (AMA) was regarded by age-mates as one who dislikes school, gets by, has to be told what to do, does what he feels, yet depends on peers for approval. Impulsivity and avoidance by peers were characteristics of Ss who were deviant and lacked "model value."

A factor of "Socially Oriented Achievement Motivation" (SOAM) was also found. This factor was defined by acceptance of school and cultural standards and by scholastic motivation. Stability, restraint, relatively little criticism of education, and a tendency to be sociable were characteristic of Ss high in this factor.

Finally, there is widespread support in the literature for the idea that high achievers are more motivated to achieve in school than are low achievers (Taylor, 1964).

In summary, high achievers tend to identify more with parents, authority figures, and with cultural norms, and they employ socially-accepted behavior. They are motivated to achieve highly in school and gain satisfaction from their school activities. They are independent and leaderlike, and enjoy peer relations without being dependent on peer group support. Low achievers, in contrast, tend to reject parental

and school values, resent authority, employ less socially-accepted behavior, are seen as deviant from the peer group, yet tend to lean on the peer group for support.

PROCEDURE

The Sample

Two Midwestern cities cooperated in this study. The larger city (population 175,000) includes a state university, is the seat of state government, is a tourist and trade center, and is the home of both heavy and light industry; all SES classes are present, although the majority of families would be considered of upper-middle or lower-middle SES. The smaller city (population 40,000) is primarily industrial in character; the population tends to lower-middle SES.

An initial sample of 310 eighth-grade junior high school boys was randomly drawn from 18 junior high schools. The Ss met the criterion of California Test of Mental Maturity IQ 120 or above, normal eighth-grade age, normal and comparable class loads, and absence of incapacitating physical defects or gross emotional disturbance. The sample was tested with the WISC, and Ss of IQ 116 or above were retained. Attrition reduced the final sample to 261 Ss.

The final sample was divided into three equal groups of high, average, and low achievers, on the basis of an equal-weight composite ranking of GPAs in the 7th and 8th grades and total score on the Iowa Test of Basic Skills. Complete data were obtained for 88 high, 85 average, and 88 low achievers.

Instrumentation

Germane to this part of the study are the student interview and card sort. A structured interview with allowance for elaboration by Ss was administered individually. Based on Ringness' (1963) previous study, the interview consisted of 55 questions tapping areas of occupational ambitions, identification with father, identification with teacher, acceptance of school values, peer relations, attitude toward school marks received, out-of-school activities, and attitude toward heterosexual relationships.

A rectangular card sort assessed dimensions of pupil self-report concerning independent behavior, nonconforming behavior, motive to affiliate with peers, and motive to achieve academically. Based on the earlier study, additional items were added to dif-

ferentiate more clearly between nonconforming and independent behavior, validity of new items being judged by trainees in School Psychology. Eighty items were provided, 20 relating to each dimension. Eight cards were sorted into each of 10 compartments, and represent scores from 1 to 10; low scores represented "most like me" and high scores represented "least like me." Coefficients of internal consistency were .51 for the dimension of independence, .71 for motive to affiliate with peers, .85 for motive to achieve academically, and .93 for nonconforming behavior, as stepped up the Spearman-Brown formula.

RESULTS

The Interview

Table 1 presents only the relevant interview data, together with chi-square tests of significance. It should be noted that because responses were sometimes elaborated, a given S's responses might be included in two or more categories so that total responses to an item did not always equal 261.

For many of the data the total responses may be more interesting than the differences between groups. For example, Ss in this study enjoy a high frequency of activities with their fathers, and strongly admire them or admire them with some reservation. Ss do not tend to admit to poor father-son relationships and there are no significant differences among achievement groups in this regard.

Although significantly more high than average or low achievers state that school is valuable for helping them develop their talents, and although significantly more average and low than high achievers stress the value of school for vocational preparation, the majority of Ss view school as primarily useful for vocational preparation. Since "vocational preparation" at the eighth-grade level is primarily that of gaining general educational tools, it would seem that (with the exception of a few high

TABLE 1
INTERVIEW RESPONSES OF BRIGHT EIGHTH-GRADE BOYS
(*N* = 88 high, 85 average, and 88 low achievers)

Interview area	Number of responses		
	HA	AA	LA
Identification with father			
Frequency of father-son activities:			
Week-ends or oftener	66	66	71
2 or 3 times per month	10	9	8
Once a month	7	3	1 ^a
Never	1	2	3
Don't know	4	6	5
Attitude toward father:			
Strong admiration	47	42	49
Admire with reservations	29	27	24
Not close	5	9	7
Desire to be different	1	0	1
Strong negative	0	2	1
Don't know	6	6	6
Identification with school			
Values of school:			
Not important	5	3	2
For occupational preparation only	58	75 ^{**a}	82 ^{**a}
To develop one's talents	17	2 ^{**a}	1 ^{**a}
Only for interest in subjects	8	8	3
Don't know	1	1	0
Values of school for future occupation:			
Very important	51	52	53
Somewhat important	18	14	11
Little or no importance	12	11	10
Don't know	6	9	14
Identification with teacher's values			
Perceived teacher norms for model pupil role:			
Conformity	79	66 ^{*a}	77
Social competence	2	14 ^{**a}	9 ^{*a}
Academic competence	22	14	14
Intellectual liveliness	1	12 ^{**a}	9 ^{*a}
Don't know	1	1	3
Attitude toward perceived norms:			
Approve	76	72	75
Indifferent	3	9	9
Dislike	2	3	3
Don't know	7	2	1
Characterization of self like the perceived model:			
Much like model	81	54 ^{**a}	61 ^{**a}
A little like model	0	9 ^{**a}	6 ^{*a}
Try to be like model	0	10 ^{**a}	3 ^{**b}
Unlike model	6	11	11
Don't know	1	1	7
Perceived teacher opinion of <i>S</i> :			
Good, <i>S</i> cares	55	40 ^{*a}	33 ^{**a}
So-so, <i>S</i> cares	19	22	29
Poor, <i>S</i> cares	6	8	11
Good, <i>S</i> doesn't care	0	1	2
So-so, <i>S</i> doesn't care	2	2	2
Poor, <i>S</i> doesn't care	0	2	3
Don't know	8	11	7

TABLE 1—Continued

Interview area	Number of responses		
	HA	AA	LA
Peer relationships			
Characteristics of popular peers:			
Athletic	14	7	8
Scholarly	5	0 ^a	3
Good personality	36	21 ^a	21 ^a
Athletic, good personality	20	25	17
Scholarly, good personality	6	9	18 ^a
Athletic, scholarly	2	5	5
All three	9	18	16
Don't know	4	0 ^a	4 ^{***}
Peer attitude toward good students:			
Admire	49	41	43
Indifferent	8	12	8
Depends on personality	15	8	12
Square (negative attitude)	13	22	23
Don't know	3	4	3
Characteristics of S's most admired peer:			
Athletic	6	8	10
Scholarly	26	12 [*]	20
Personality	31	20	25
Athletic, personality	3	3	5
Scholarly, personality	15	10	12
Athletic, scholarly	0	0	0
All three	29	22	20
Social status	0	1	0
No admired peer	2	6	4
Don't know	5	5	6
Peer attitude toward school tasks:			
Work hard as possible	12	13	9
Work little more than is necessary	18	21	15
Work enough to get by	55	47	61 ^{***}
Don't know	3	5	3

Note.—All chi-square tests of significance.

^a Comparison with high achievers.

^b Comparison with average achievers.

* $p < .05$.

** $p < .01$.

achievers) school is more of a routinely-accepted environment than a place which is viewed as important to self-development or which provides interesting content and activities. This, in turn, may be related to the perceived teacher norms for model pupil behavior.

Most Ss state that teachers view the "model pupil" as conforming. For the sample as a whole, academic and social competence and intellectual

liveliness fare less well. Although there are statistically significant differences in the ways high achievers differ from average and low achievers as to perceived norms of social competency and intellectual liveliness, it is noteworthy that there is high agreement concerning conformity. By implication, achievers must be conformers, a finding which agrees with those of studies cited earlier.

Differences in achievement are not

TABLE 2
CARD SORT DATA FOR BRIGHT EIGHTH-GRADE BOYS
(N = 88 high, 85 average, and 88 low achievers)

Dimension	Scores		Mean differences		
	$\bar{X}$	SD	Groups	Difference	t
Nonconformity:					
High achievers	7.30	.74	H-A	-.02	.17
Average achievers	7.32	.82	H-L	.40	3.64*
Low achievers	6.90	.73	A-L	.42	3.50*
Independence:					
High achievers	4.98	.64	H-A	-.08	.80
Average achievers	5.06	.68	H-L	-.26	2.60*
Low achievers	5.24	.69	A-L	-.18	1.80
Achievement motivation:					
High achievers	4.17	.77	H-A	-.32	2.67*
Average achievers	4.49	.80	H-L	-.81	6.23*
Low achievers	4.98	.90	A-L	-.49	3.77*
Affiliation motivation:					
High achievers	5.15	.96	H-A	.46	3.07*
Average achievers	4.69	1.04	H-L	.60	3.33*
Low achievers	4.55	.87	A-L	.14	.93

* $p < .01$.

related to differences in pupil perceptions of teacher expectations, but they are related to attainment of the model pupil role. Although most Ss state that they approve the conformity role, high achievers feel that they are like the model more than do average or low achievers; average and low achievers state significantly more often that they are a little like the model, and average achievers mention that they "try to be" like the model. There are no significant differences in the extent to which achievement groups state that they are unlike the model. Most Ss say that they care how they are perceived by the teacher.

Taken as a whole, these data suggest that most Ss identify well with their fathers and with teachers. They perceive school as a place one attends for vocational purposes, and where conformity is demanded. At this point in their lives they are generally willing to accept the conformity role, and care how they are regarded by their teachers.

In regard to peer relationships it is patent that having a "nice personality" is the most important factor in popularity. Good students are admired by about half the sample, but 35 Ss mention that a good student is popular primarily if he has appropriate personality attributes. Fifty-eight Ss, or about one-fifth of the sample, describe good students as "square" and another 28 are indifferent to scholarship attributes. These findings are also borne out in responses to questions concerning characteristics of most-admired peers. The majority of Ss feel that the effort norm of the peer group is that of doing just enough to "get by," although 54 think most peers do a little more than necessary.

Thus although school norms are perceived as those of conformity and Ss accept this norm and attempt to be like the perceived model, the model is not viewed by the majority as that of intellectual development so much as that of "good behavior." This seems related to the responses which indicate

that peer achievement norms are essentially those of "getting by" or doing "just a little more than necessary." Popularity seems to result from possession of a pleasing personality; athletic prowess, scholarship, and other attributes may supplement such personal qualities, but are in themselves not conducive to peer acceptance.

The Card Sort

Table 2 presents card sort data, together with two-tailed tests of significance of differences. It will be recalled that scores might range from 1 to 10, with lower scores being "most like me."

It is seen that high and average achievers are less nonconforming or less rebellious than low achievers. On the other hand, high achievers are more independent and autonomous than low achievers.

High achievers are significantly more oriented toward academic achievement than average achievers, who are significantly more oriented toward achievement than are low achievers. But average and low achievers are significantly more oriented to affiliate with peers than are high achievers.

DISCUSSION

This study was undertaken partly to confirm or refute findings of an earlier study. In most respects but not all, confirmation was found. Within the limits of changes in sampling and instrumentation mentioned earlier, some comparison of findings may be made.

In both studies the father was admired, or admired with some reservations, by a majority of Ss; however, the 1963 findings showed that fathers of high achievers spent more time with their sons than did fathers of low

achievers, suggesting closer father identification, but this was not found in the present study. It cannot be concluded from the present (and larger) study that father identification differentiates between high, average, and low achieving Ss. There is one possibility, which was not attacked in either study, which might account for the differential findings. It was noted earlier that the prior study dealt with a sample from the upper-middle SES population whereas the present sample dealt with all SES groups. If, as was suggested, upper-middle SES fathers are more oriented toward high scholastic achievement than fathers of other SES groups, differences in the two studies may simply represent differences in parental value-orientations. This tends to confirm McGuire's notion that talent factors are possibly community-bound. At the moment this must be regarded as an area for future research.

In regard to identification with the teacher or with school values, both studies showed Ss to perceive the teacher's concept of the model pupil as one who conforms, or poses no problems in the classroom. In the 1963 study high achievers approved this role concept but low achievers did not; the present data show no significant differences among groups. However, both studies show that high achievers feel that they are more like the role model than are other groups; although most Ss desire to be viewed as like the role model.

Data in both investigations showed that school is viewed by most students as valuable for vocational preparation; however, high achievers more than others saw school as a place to develop one's talents. "Good students" were accepted by about half the Ss in each study; the 1963 data showed that

high achievers admired scholars more than did low achievers, but this difference was not found in present data. Good students are characterized as "squares" by about one-fifth of the present sample; in previous data low achievers were more likely to make this statement. In both studies the peer norm for academic achievement was that of doing just enough to "get by."

Both studies showed that high achievers are motivated to achieve academically more than other groups, whereas low achievers are more motivated to affiliate with peers. High achievers are presently found to be independent, whereas low achievers are somewhat nonconforming; this finding appears to document statements of Taylor (1964).

The work of McGuire et al. isolated certain predictors of talent which were noncognitive. Among these were Peer Stimulus Value, Age-Mate Avoidance, and Socially Oriented Motivation. (They also found male-specific factors of Anxious Emotionality and Antisocial Wariness, but these were less clearly specified.) PSV suggests that the talented are likely to be leaderlike, independent, and effective. This is similar to present findings in the card sort that high achievers are more independent than other Ss. PSV, however, suggests that possessors are models for peers. Interview findings of the two studies by Ringness attacked the questions of popularity and "most-admired peers" and it was found that models for these samples were chosen more for "personality" than any other reason.

The AMA factor is similar to card sort findings that low achievers are more nonconforming, yet more affiliation-oriented than high achievers. The SOAM factor is essentially confirmed by both card sort and interview find-

ings that high achievers are motivated to conform to school and adult values, they identify well with home and school, and are academically motivated.

The main difference in the findings of the McGuire and the present study is in the implications in the former that peers tend to admire the talented, whereas in present findings popularity and admiration are associated with achievement primarily by high achievers, when so associated at all. In other respects, although the factors combined attributes of achievers in a somewhat different fashion from the present study, essentially the same characteristics were found; these are also similar to those summarized in Taylor's review.

There seems sufficient evidence to support the conclusion that low achievers identify more with the peer group and are governed more by the peer group than are high achievers. Since the peer norm for achievement is seen as that of mediocrity, and since low achievers state less motivation for academic achievement, it would follow that an important problem to schools is that of finding ways to foster high achievement values in the peer group at large. The image of the scholar as "square" must be erased. It is possible that this image is fostered partly by the fact that a certain percentage of high achievers are not well-rounded persons, lack desirable personality attributes, and are not popular.

The image of the teacher needs revision. He is viewed as demanding conforming behavior, but is not seen as fostering intellectual development or liveliness. Schools are seen as places to prepare for future vocations, but are seen by relatively few as places to develop talents, pursue interest, or to improve social adjustment.

The present study refutes the no-

tion that high achievers identify more with their fathers than do low achievers. However, at junior high school age, parent identification apparently does not provide reinforcements for academic achievement to low achievers to the same extent it does for high achievers. It is possible that differential parent values are operating, and it would seem desirable for school counselors to discuss such matters with fathers of low achievers when the attainments of the latter are of sufficient concern.

One other comment may be justified. It is apparent that the stereotype of school tasks as being necessary evils seems still to exist. Social reinforcements such as teacher praise and blame, grades and marks, and other typical reinforcements provided by the school are not as effective with low achievers as could be wished. The implication is that much more effort needs to be spent on finding other kinds of reinforcers for academic achievement. Efforts to employ concrete reinforcers may provide some clues; efforts to make the curriculum more meaningful to boys like those in these samples may also bear fruit.

REFERENCES

- BANDURA, A., & WALTERS, R. H. *Adolescent aggression*. New York: Ronald Press, 1959.
- BANDURA, A., & WALTERS, R. H. *Social learning and personality development*. New York: Holt, Rinehart and Winston, 1963.
- DREWS, E., & TEAHAN, J. Parental attitudes and academic achievement. *Journal of Clinical Psychology*, 1957, **13**, 328-332.
- FRANKEL, E. A comparative study of achieving and underachieving high school boys of high intellectual ability. *Journal of Educational Research*, 1960, **53**, 172-179.
- GOWAN, J. D. The underachieving gifted child. *Exceptional Children*, 1955, **21**, 247-249.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, L. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- MCGUIRE, C., HINDSMAN, E., KING, F. J., & JENNINGS, E. Dimensions of talented behavior. *Educational and Psychological Measurement*, 1961, **21**, 3-38.
- MORROW, W. R., & WILSON, R. R. Family relations of bright high-achieving and under-achieving high school boys. *Child Development*, 1961, **32**, 501-510.
- PORTLAND PUBLIC SCHOOLS. *The gifted child in Portland*. Portland, Oregon, 1959.
- RINGNESS, T. A. Differences in attitudes toward self and others of academically successful and non-successful ninth-grade boys of superior intelligence. Final report, post-doctoral research fellowship sponsored by National Institute of Mental Health. University of Wisconsin, Madison, 1963. (Mimeo)
- RINGNESS, T. A. Affective differences between successful and non-successful bright ninth-grade boys. *Personnel and Guidance Journal*, 1965, **43**, 600-606. (a)
- RINGNESS, T. A. Emotional adjustment of academically successful and non-successful bright ninth-grade boys. *Journal of Educational Research*, 1965, **59**, 88-91. (b)
- RINGNESS, T. A. Non-intellective variables related to academic achievement of bright junior high school boys. Final report, Co-operative Research Project S-036, United States Office of Education. Madison, Wisconsin, 1965. (c) (Mimeo)
- TAYLOR, R. G. Personality traits and discrepant achievement: A review. *Journal of Counseling Psychology*, 1964, **11**, 76-82.
- THORNDIKE, R. L. *The concepts of over- and underachievement*. New York: Bureau of Publications, Teachers College, Columbia University, 1963.

(Received May 11, 1966)

EFFECTS OF STRATEGY, SEX, AND AGE ON CONCEPTUAL BEHAVIOR OF ELEMENTARY SCHOOL CHILDREN

GLENN E. TAGATZ¹

Indiana State University

The effects of strategy, sex, and age on 4 concept-attainment problems were examined. Strategies were a commonality and a conservative method of solution. The age and sex dichotomies resulted from 5th and 6th grades. Time-to-criterion scores were the dependent variable. Departure from instructional strategy was examined using chi-square. Statistically significant differences were found for all main effects and the repeated measures in the first analysis. The strategy difference favored the commonality solution. The sex difference favored the females. The 6th grade was less efficient than the 5th grade, indicating greater awareness of the complexity of the task and of the stimuli. The chi-square analysis indicated statistically significant preference among all Ss for the commonality strategy.

Researchers, Bruner, Goodnow, and Austin (1956), Klausmeier, Harris, and Wiersma (1964) and Miller, Galanter, and Pribram (1963), have attempted to identify strategies used by subjects (Ss) in the attainment of concepts. Following behavioristic tradition these efforts have largely consisted of examination of instances selected (selection strategies) or the relationship between instances and subsequent behavior (reception strategies). The concept of strategy as a plan allows another approach which has, until recently, been overlooked in attempts to identify strategies. This consists of (a) "programming" Ss with the plan or strategy, (b) having Ss attain a concept, and (c) then making subsequent examination to determine whether or not the respective strategy has been followed and to determine the relative efficiency of Ss using the strategies studied. The objectivity of earlier efforts is maintained; and additionally, the relative efficacy of the strategies compared may be appraised.

This particular method of studying strategy has the additional advantage of eliminating some of the residual variability in an analysis which arises from the fact that Ss previously operated initially without a clear frame of reference as to any of the possible methods of attainment. Hence early attempts by Ss were characterized by inefficiency regarding knowledge of method in addition to inefficiency regarding knowledge of materials, that is, levels of attributes, number of attributes relevant to the concepts in the study, etc.

The influence of the sex of an S on concept attainment has been considered by various researchers with varying results; Klausmeier et al. (1964), Tagatz and Meinke (1966), and Olson (1963). In the first two studies, students of college age were used. In Olson's study adolescents were used. The first two failed to find a statistically significant sex effect; the last did. Olson concluded that the demonstration of sex differences depends upon the particular concept to be learned. It may depend also upon development and concomitant characteristics.

¹ The author is presently a Postdoctoral Fellow at the Research and Development Center for Learning and Re-Education at the University of Wisconsin.

The age of Ss, then, is another variable which must be given attention in conjunction with the study of concept attainment. Yudin and Kates (1963) found that 14- and 16-year-old Ss were significantly superior to the 12-year-old Ss in their study. They concluded that their results confirm the description of adolescent development advanced by Inhelder and Piaget (1958), who suggest that formal reasoning begins at about age 12 years and reaches equilibrium at about 14-16 years of age. In the present study, Ss were selected from fifth and sixth grades in an attempt to examine performance immediately prior to and at the age at which such formal reasoning begins. If performance on tasks used in this study reflects formal reasoning, then differences between the two grades should be evident.

THE PROBLEM

The purpose of this study, then, was to investigate the influence of certain variables upon efficiency of concept attainment. Specific questions examined were:

1. What are the effects of the following four variables upon concept attainment: (a) grade level—fifth or sixth; (b) sex; (c) instructed method of solution—conservative or commonality strategy; and (d) repeated measures, that is, four problems?

2. Is there a significant preference among fifth- and sixth-grade students for either one of the two strategies?

In this study the conservative strategy consisted of a verbally presented algorithm for concept attainment in which information processing of one-dimensional differences was specified. For the commonality strategy, a method of solution was articulated whereby the defining characteristics common to a set of conceptual exemplars were specified as the concept,

provided complete dimensional variation was presented.

METHOD

Subjects

The Ss for this study consisted of a stratified, random sample of fifth- and sixth-grade students from the Campus Elementary School at Oshkosh, Wisconsin. From each grade level 10 males and 10 females were included. Thus a total of 40 students participated.

Material

Stimulus material for the study consisted of a display board containing 32 cards. Each card contained one of two levels of figurally presented information for each of five dimensions. The five dimensions and their defining characteristics (levels) follow: (a) border number: one or two; (b) border continuity: solid or broken; (c) figure number: one or two; (d) figure color: purple or yellow; (e) figure shape: square or triangle.

On every card one level of each of the five dimensions was presented. Each card was different from every other card in one or more of the five possible ways. This stimulus construction permitted conjunctive categorization of the cards on the board into groups possessing common characteristics and those not possessing the combination of characteristics.

Design and Procedure

Upon arrival at the learning laboratory Ss were told the general nature of the experiment and then received one of two sets of instructions. Each set had in common general directions about the conduct of the experiment and special instructions about a possible method of solution.

In the general directions, each pupil was told that the experiment was concerned with how people learn and that he would be given an opportunity to work several problems. The stimulus construction was explained and repeated until Ss were well aware of the five dimensions. This degree of familiarity was assured by having students describe each of the five dimensions as the next step in the general directions. Pupils were made aware of the way that cards could be grouped into those categories possessing common characteristics and those not possessing the characteristics. Thus, a concept was defined for the students.

It was additionally explained that a concept was merely a rule for classification. Understanding of categorization was determined by having Ss specify which cards belonged to a category articulated by the examiner; for example, Ss were asked to identify which cards belonged to the conceptual category "two borders with purple figures." A verbal response of "yes" was given to cards which illustrated the rule and a verbal response of "no" was given to cards which did not. The Ss were then informed that their job was to determine the rule that the examiner had in mind when "yes" and "no" markers were arranged on the cards constituting the stimulus display. At this particular point, the special instructions were given to each respective group. The instructions for the commonality and the conservative strategies follow:

Instructions for Commonality Strategy:

You will be able to attain the concept successfully by giving your entire attention to the "yes" cards. To determine what is the correct concept, examine the dimensions of the "yes" cards. Those dimensions which are common to all "yes" cards specify the relevant features of the concept. As an example, consider the following three cards: (Figural dimensions are verbally presented.)

Card 1 (yes)

1. One Border
2. Solid Border
3. Two Figures
4. Purple Figures
5. Triangular Figures

Card 2 (yes)

1. Two Borders
2. Solid Borders
3. Two Figures
4. Yellow Figures
5. Square Figures

Card 3 (no)

1. One Border
2. Broken Border
3. Two Figures
4. Purple Figures
5. Square Figures

The markings indicate that the first two cards belong to a group of cards which exemplify a concept; the third does not. Both "yes" cards possess the dimensions of two figures and solid borders. They are different in the other dimensions. The concept or underlying rule might be, therefore, (a) "two figures," (b) "solid borders," or (c) "two figures and solid borders." By examining other "yes" cards, you could determine exactly which of the

three possibilities is correct. If some other "yes" card contained one figure, the only correct concept would then be "solid borders." If all the "yes" cards contained two figures and solid borders, "two figures and solid borders" would be the concept.

Instructions for Conservative Strategy:

You will be able to attain the concept successfully by examining cards which do and do not belong to the group of cards which exemplify the concept. Specifically, a "yes" card which differs from a "no" card in one dimension tells you one relevant or important dimension of the concept. As an example, consider the following three cards: (Figural dimensions are verbally presented.)

Card 1 (no)

1. Two Borders
2. Solid Borders
3. Two Figures
4. Purple Figures
5. Square Figures

Card 2 (yes)

1. Two Borders
2. Broken Borders
3. Two Figures
4. Purple Figures
5. Square Figures

Card 3 (?)

1. Two Borders
2. Solid Borders
3. Two Figures
4. Yellow Figures
5. Square Figures

The center card [Card 2] illustrated the concept. The card on the left [Card 1] does not illustrate the concept. These two cards are different in only one dimension, namely, border continuity. Therefore, cards which illustrate the concept must have broken borders. Would the card on the right [Card 3] illustrate the concept?

You should have answered this question with a "no." The card on the right has solid borders. We know from the other two cards that cards which illustrate the concept must have broken borders. The third card then cannot illustrate the concept. At this point you would know one of the relevant features of the concept. By examining other "yes" and "no" card combinations, you can determine exactly what is the correct concept.

Following the special instruction all Ss were shown how to respond with their concept. This was done by using a slip of paper on which the five dimensions and their two levels were specified. The Ss' task was to check the relevant level of dimensions

which specified the concept. Upon completion of these instructions each *S* was given an opportunity to ask questions about anything which was not clear. When questions did arise, the appropriate section of the instructions was repeated and explained.

In the experiment *Ss* were instructed to solve four problems. Problems were conjunctive with two dimensions relevant for each. Two exemplars of the concept—one the exemplar focus card—and two non-exemplars were specified for each problem. The exemplar which shared with the focus card the two relevant dimensions was indicated with a "yes" marker. Two nonexemplars, which differed from the focus card in only one relevant dimension each, were specified with "no" markers. Hence, equal amounts (bits) of exemplar and nonexemplar information were presented regarding each concept.

Data of two types were recorded by the examiner; namely, (a) time taken to attain successfully the concept, and (b) which cards *S* used in the attainment of the concept. This last variable was secured at the completion of all four problems. By knowing which cards *S* used in the attainment of the concept, it was possible to determine whether he had used the respective strategy set forth in the special instructions given him. Upon completion of the fourth problem, then, *Ss* were asked which cards they had used in the attainment of the concept. It was expected that *Ss* who received the commonality instructions would indicate use of the exemplars and that *Ss* who received the conservative instructions would use the nonexemplars. Departure from this conformity to strategy was appropriately noted by the experimenter for later analysis.

ANALYSIS

Time-to-criterion scores were examined in a $2 \times 2 \times 2$ factorial, repeated measures design. Stratifying variables were grade—fifth and sixth—and sex. The treatment variable consisted of a comparison of the conservative and commonality strategies. These three main effects and their interactions constituted the first part of the analysis. The second part was the examination of repeated measures, that is, the four problems. The third

and last part of the analysis consisted of the interaction of the repeated measures with the main effects and their interactions.

The second dependent variable—departure from instructional mode (strategy)—was analyzed using chi-square techniques. In this analysis the frequency of departure from the instructional method of solution was collapsed over stratifying variables leaving a 2×2 table, namely, commonality versus conservative instructions and commonality versus conservative solutions. Thus the relationship of instruction to method of solution was examined.

RESULTS

Table 1 presents the results of the analysis of variance performed on

TABLE 1
ANALYSIS OF VARIANCE ON TIME-
TO-CRITERION SCORES FOR
CONCEPT-ATTAINMENT
PROBLEMS

Source	df	MS	F
Grade (G)	1	9.79	4.20*
Sex (S)	1	11.66	5.03*
Instructional Strategy (I)	1	36.66	15.80**
G $\times$ S	1	.51	—
G $\times$ I	1	6.26	2.69
S $\times$ I	1	12.11	5.22*
G $\times$ S $\times$ I	1	.61	—
S/Treatments	32	2.32	—
Repeated Measures	3	31.66	20.69**
G $\times$ RM	3	5.03	3.29*
S $\times$ RM	3	.99	—
I $\times$ RM	3	14.36	9.39**
G $\times$ S $\times$ RM	3	.27	—
G $\times$ I $\times$ RM	3	8.06	5.27**
S $\times$ I $\times$ RM	3	1.43	—
G $\times$ S $\times$ I $\times$ RM	3	.69	—
S/T $\times$ RM	96	1.53	—
X	1	—	—
Total	160	—	—

* $p < .05$.

** $p < .01$.

time-to-criterion scores.² Tables 2 through 6 present mean times-to-criterion for selected significant effects in Table 1 and their respective Duncan's range.

Of the three main effects and their interactions, four sources of variance were significant. The Ss that had received the commonality instructions were more efficient than those that had received the conservative instructions. The mean for the Commonality Instructional Group was .90 minutes, for the Conservative Instructional Group 1.86. The sixth graders demonstrated less efficient solution than did the fifth graders, the mean for the

TABLE 2
MEAN TIMES-TO-CRITERION FOR SEX
× STRATEGY INTERACTION AND
DUNCAN'S RANGE

Subjects	Con- servative instructions	Com- monality instructions	Duncan's range
Males	2.41	.90	.8
Females	1.32	.91	

fifth graders being 1.13 minutes, for the sixth graders 1.63 minutes. This difference was significant at the .05 level. The females were more efficient than the males; the mean time for the females was 1.11 minutes, for the males 1.65 minutes. This difference was significant at the .05 level. Table 2, the Sex × Instructional Strategy interaction, indicates that males receiving conservative instructions were significantly less efficient than the females receiving that set of instructions as well as both males and females receiving commonality instructions.

² Levels of significance of *F*-ratios involving repeated measures reflect consideration of Greenhouse and Geisser (1959).

TABLE 3
MEAN TIMES-TO-CRITERION FOR
REPEATED MEASURES AND
DUNCAN'S RANGE

Ordinal position				Duncan's range
One	Two	Three	Four	
2.66	1.27	.91	.68	1.9

Table 3 presents the means and Duncan's range for the four ordinal positions. The difference between Ordinal Position 1 and Ordinal Position 4 was significant. The remaining differences were not significant. The fact that the first problem was especially difficult for Ss attempting to use the conservative strategy is shown in Table 4. None of the other differences were significantly different in this Instruction × Repeated Measures interaction.

The significant Grade × Repeated Measures interaction, the means for which are reported in Table 5, resulted because the performance of the sixth-grade Ss was significantly different on the first problem than the third and fourth problems and from all four problems for the fifth-grade Ss. A similar phenomenon is evident in Table 6 where the Grade × Instruction × Repeated Measures interaction reveals Ordinal Position 1 to be especially difficult for sixth-grade Ss who have received the conservative

TABLE 4
MEAN TIMES-TO-CRITERION FOR STRATEGY
× REPEATED MEASURES INTERACTION
AND DUNCAN'S RANGE

Instruction	Ordinal position				Duncan's range
	One	Two	Three	Four	
Conservative	4.04	1.45	1.14	.82	2.5
Commonality	1.29	1.10	.67	.55	

TABLE 5

MEAN TIMES-TO-CRITERION FOR GRADE $\times$
 REPEATED MEASURES INTERACTION AND
 DUNCAN'S RANGE

Grade	Ordinal position				Duncan's range
	One	Two	Three	Four	
Fifth	1.89	1.14	.88	.63	2.5
Sixth	3.44	1.41	.94	.74	

instructions. Close examination reveals other significant ranges for which psychological implications are not readily apparent.

The chi-square analysis of Ss' departures from instructional strategy

TABLE 6

MEAN TIMES-TO-CRITERION FOR GRADE $\times$
 STRATEGY $\times$ REPEATED MEASURES
 INTERACTION AND DUNCAN'S RANGE

Instructional strategy	Grade	Ordinal position				Duncan's range
		One	Two	Three	Four	
Conservative	Fifth	2.40	1.41	1.06	.79	1.3
	Sixth	5.68	1.48	1.22	.84	
Commonality	Fifth	1.38	.87	.69	.47	
	Sixth	1.20	1.33	.65	.63	

was statistically significant at the .001 level. Table 7 presents the respective values of this analysis. It is evident from this analysis that there is a distinct preference for the commonality strategy when it is contrasted with the conservative strategy and when Ss are elementary school children from

the fifth and sixth grades. Six fifth graders and four sixth graders of the 10 from each grade which received the conservative instructions reverted to the commonality solution. No Ss switched from the commonality instructional strategy to the conservative solution.

DISCUSSION

Two distinct strategies, both of which lead to the attainment of concepts, have been contrasted. The significant difference between them indicates that most of the fifth and sixth grade Ss used in the study were not able to use the conservative strategy. Of those that did, all required above the median time-to-criterion over all problems revealing the conservative strategy to be less efficient for young Ss in the intermediate grades. The nature of the conservative strategy requires greater use of formal logic than does the commonality plan; and hence, this finding adds empirical support to Inhelder and Piaget (1958). Additionally supportive is the fact that the performance of sixth-grade was inferior to that of fifth-grade students. It seems likely that the sixth graders as a group became aware of the greater complexity of the task and the combinatorial aspects of stimuli than did the fifth graders. This is further supported by the significant Grade $\times$ Instruction $\times$ Repeated Measures interaction in which the sixth graders who had re-

TABLE 7

CHI-SQUARE OF DEPARTURE FROM INSTRUCTIONAL STRATEGY

	Commonality solution		Conservative solution	
	$\frac{(20 - 15)^2}{15} = 1.66$		$\frac{(0 - 5)^2}{5} = 5$	
Commonality instruction	$\frac{(10 - 15)^2}{15} = 1.66$		$\frac{(10 - 5)^2}{5} = 5$	
Conservative instruction	30		10	
			40 $\chi^2 = 13.32$	

ceived the conservative instructions were significantly less efficient initially than all others. It appears that the fifth graders attained the concept by looking at the common elements (class inclusion) and just missed or ignored anything else—unless their attention was drawn specifically to it—thus demonstrating the absence of concern for finer discrimination of elements which is present in more formal thought. The significant Grade $\times$ Repeated Measures interaction also supports this conclusion as does the Instruction $\times$ Repeated Measures interaction. More efficient behavior occurred as Ss reverted to Piaget's class inclusion operation at the concrete level.

While not a unique finding of this study, the significant sex effect does reflect organismic differences in conceptual behavior. It is doubtful that this significance resulted because of a Sex $\times$ Concept interaction as was Olson's conclusion. Rather it is likely that the difference resulted because of advanced verbal development of females. The significant Sex $\times$ Instruction interaction also supports this conclusion. Males would generally be expected to perform in a fashion superior to females on an analytic task. As this is not the case, the verbal superiority explanation seems most tenable.

Also of major significance is the empirical validation that "strategy" can be effectively studied through a three stage process, namely (a) "programming" Ss with the plan or strategy, (b) having Ss attain a concept, and (c) then making subsequent examination to determine whether or not the respective strategy has been followed and to determine the relative efficiency of Ss instructed to use the

strategies studied. Results of the present study reflect different developmental levels in children's thinking. The reduced efficiency of the sixth-grade Ss reflected initial preoccupation with a combinatorial system (formal thought) not present in the fifth-grade Ss. This apparently inherent and increased mediational activity in the older Ss seemed to be reinforced by instructions regarding a conservative strategy though many Ss so "programmed" eventually reverted to a class-inclusion operation. Thus it was shown that the three-stage process can be effectively used to study conceptual behavior even to the extent of comparing "programmed strategies" interacting with other variables.

REFERENCES

- BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
- GREENHOUSE, S. W., & GEISSER, S. On methods in the analysis of profile data. *Psychometrika*, 1959, **24**, 95-112.
- INHELDER, B., & PIAGET, J. *The growth of logical thinking*. New York: Basic Books, 1958.
- KLAUSMEIER, H. J., HARRIS, C. W., & WIERSMA, W. *Strategies of learning and efficiency of concept attainment by individuals and groups*. University of Wisconsin: Cooperative Research Project No. 1442, 1964.
- MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior*. New York: Holt, 1960.
- OLSON, L. A. Concept attainment of high school sophomores. *Journal of Educational Psychology*, 1963, **54**, 213-216.
- TAGATZ, G. E., & MEINKE, D. L. Information processing in concept attainment tasks. *Teachers College Journal*, 1966, **37**, 182-186.
- YUDIN, L., & KATES, S. L. Concept attainment and adolescent development. *Journal of Educational Psychology*, 1963, **54**, 177-182.

(Received May 19, 1966)

IMPLICIT ASSOCIATIVE RESPONSE OCCURRENCE IN LEARNING WITH RETARDED SUBJECTS:

A SUPPLEMENTARY REPORT

WILLIAM P. WALLACE¹

Northwestern University

100 words were read to groups of normal and retarded Ss. As each word was read Ss were required to indicate whether it had been read earlier in the list. Words appeared in the list which were presumed to elicit specific implicit associative responses (IARs), and later in the list the presumed IARs were presented. A significant interaction indicated that this manipulation resulted in a greater increase in error rates for normal Ss than for retarded Ss. The results were interpreted as supporting the notion that retarded Ss make fewer or weaker IARs when presented with verbal units.

The classification of mental retardation implies a learning deficiency. An understanding of the nature of this learning deficiency should be of practical importance. When specific deficiencies are understood, classroom training can proceed most efficiently, compensating for weaknesses of the students and capitalizing on their strengths. The present study was designed to investigate one potential source of deficiency in learning of retarded subjects (Ss), namely the occurrence of implicit associative responses.

The purpose of the present study was to investigate the influence of implicit verbal responses in producing recognition errors among normal and retarded Ss. A familiar word presented as a stimulus in a verbal-learning task may elicit two types of implicit responses. Bousfield, Whitmarsh, and Danick (1958) have labeled as the "representational response" (RR) the response to the word itself as the act of perceiving it. The second kind of implicit response, the implicit associative response (IAR), is produced by the stimulus properties of the RR and is conceived of as being related to the RR. For example, as the act of

perceiving the letters *DOG* an RR of "dog" is presumed to occur. Implicit associative responses are presumed to occur to the RR, hence "cat" is a presumed IAR to the RR "dog". It is assumed that specific IARs may be predicted from a knowledge of pre-experimental associative habits. The influence of IARs in producing recognition errors was of interest in the present study.

The task was similar to one used by Underwood (1965). The Ss were presented with a series of words and were asked to indicate whether each successive term was "old" or "new," that is, whether it had occurred before. In the list were critical words assumed to elicit particular IARs. At a later point in the list the words assumed to be IARs (experimental words) to the earlier RRs were presented. Underwood found there was a significantly greater tendency for Ss to respond incorrectly to experimental words than to control words (words presumed not to have occurred earlier as IARs). It was *as if* the experimental words had occurred earlier in the list. That is, Ss were more likely to recognize falsely *cat* when *dog* appeared earlier in the list as compared to *table* when *chair* had not appeared earlier in the list.

¹ Now at the University of Nevada.

Wallace and Underwood (1964) presented evidence consistent with the position that retarded Ss respond to verbal units with fewer or weaker IARs. Their results indicated that with college Ss IARs to conceptually related words facilitated free learning and interfered with paired-associate learning. With retarded Ss, manipulations of conceptual similarity had little effect upon free learning and paired-associate learning, resulting in a triple interaction between Class of $S \times$ Type of Task $\times$ Similarity. It was suggested that the triple interaction was due to a weaker or less frequent IAR occurrence among retarded Ss as compared to normal Ss.

The present study was designed to investigate the influence of implicit verbal responses in producing recognition errors among normal and retarded Ss. Underwood (1965) demonstrated that error rates were higher for specific words that were presumed to have occurred earlier in the list as IARs, when compared to specific words presumed not to have occurred earlier as IARs. Wallace and Underwood presented evidence suggesting that there is an IAR deficiency with

retardates. It follows that retardates should exhibit less difference than normals in error rates between words presumed to have occurred as IARs earlier in a list and words presumed not to have occurred as IARs earlier. Thus, an interaction between Class of $S \times$ Type of Word is predicted.

METHOD

Materials

The materials were taken from the Underwood (1965) recognition list of 200 words. From this list 100 words were selected to be used in the present study. The words used may be separated into four categories according to their assumed relation to other list words. Comprising one category were critical stimulus (CS) words. These words were assumed to elicit specific IARs. Experimental (E) words constituted a second category. These words represented the assumed IARs to the CS words. The third class consisted of control (C) words. These words served as controls for the E words and were presumed not to have been preceded in the list by words for which they were IARs. The final category contained filler (F) words. These words were presumed to be neutral with regard to E words and were used to establish specific repetition frequencies.

The list consisted of 15 E words, 15 C words, 44 CS words, and 26 F words. There were 14 repeated words, that is, words oc-

TABLE 1
CRITICAL STIMULUS WORDS, EXPERIMENTAL WORDS, AND CONTROL WORDS

Class	CS words	E words	Position	C words
A ₁	BOTTOM	Top	57	Down
	GIVE	Take	67	Good
	DAY	Night	85	Low
	MAN	Woman	96	Rich
A ₂	ROUGH	Smooth	77	Weak
	FALSE	True	80	Dirty
	HARD	Soft	89	Short
	SLOW	Fast	97	Girl
CV	BUTTER, CRUMB	Bread	64	Bridge
	SUGAR, BITTER, CANDY	Sweet	74	Salt
	WARM, CHILL, FREEZE, FRIGID, HOT, ICE	Cold	76	Cloud
	DARK, HEAVY, LAMP, MATCH	Light	88	Leg
SO	PEACH, GRAPE, APPLE, PEAR	Fruit	79	Cloth
	ROBIN, SPARROW, BLUEJAY, CANARY	Bird	95	Flower
SI	BANDAGE, CHALK, MILK, RICE, SNOW	White	92	Red

curing for the second or third time. The E and C words were restricted to the second half of the list. Four different relations were represented by the E words and their corresponding CS words. The E word was either an antonym (A_1 if the CS word appeared once prior to its E word and A_2 if the CS word appeared three times prior to its E word), a converging word associate (CV), a superordinate (SO), or a sense impression (SI) of its corresponding CS word(s). Some E words had more than one CS word, and some CS words were repeated more than once prior to the appearance of their respective E word. The specific CS words, E words, and C words and the list positions of the E words are presented in Table 1. A given C word was always two positions removed from its corresponding E word: two positions preceding the E word on eight occasions and two positions following the E word on seven occasions.

Procedure

Each *S* was given three sheets of paper with numbers, each number followed by the words *yes* and *no*. The first page was numbered from 1 to 10, the second page from 1 to 50, and the third page from 51 to 100. The *Ss* were instructed that they would hear a list of words, and as they heard each word, they were to ask themselves if the word had appeared in the list before. If the word occurred for the first time they were to circle *no*, but if it had appeared in the list earlier, they were to circle *yes*. A 10-word practice list was read by *E* to aid *Ss* in understanding the procedure. After the *Ss* corrected their practice list, the 100-word test list was presented by a magnetic tape recorder at a 10-second rate (10 seconds elapsed between the reading of each word). Each word was spoken twice in immediate succession, and the word number was indicated before every word. After the fiftieth word there was an additional 10-second pause, during which *Ss* were instructed to turn to the third page.

Subjects

The normal *Ss* were 31 introductory psychology students (13 males and 18 females) attending the summer session at Northwestern University. This phase of the experiment was conducted in one group session.

The retarded *Ss* were 76 residents of Dixon State School.² One problem present

in this study was a differential mortality rate between normals and retardates. The task proved to be difficult for the retardates as 36 of the original 76 *Ss* failed to comprehend the instructions, thus, the analyses were based upon the responses of 40 retarded *Ss* (25 males and 15 females). The 40 retarded *Ss* ranged in IQ from 39 to 80 with a mean of 60.5, and they ranged in age from 12.0 to 19.6 years with a mean of 16.1. The experiment was conducted in eight group sessions with group size varying from 6 to 15 *Ss*.

The *Ss* were discarded when there were indications that they did not understand the nature of the task. In most cases the decisions were unequivocal. For example, *Ss* were discarded if they failed to respond, if they circled both alternatives, or if they gave stereotyped responses (e.g., all *yes* responses). However, there was one criterion for discarding *Ss* which deserves further comment. By looking at an *S's* answer sheet it was possible to determine how many times he responded correctly, that is, circled *yes* when a word was repeated and circled *no* when a word appeared for the first time. The *Ss* responding at or below the chance level, that is, obtaining 50 or fewer correct responses, were discarded. There were 10 such *Ss*, all from the retarded sample. The problem arises because it can not be stated with certainty whether these *Ss* did not understand the task or were just poor performers. If they were included in the analysis and some of them did not understand the task, then there would be an artificial reduction in the difference between E word and C error rates among the retarded *Ss*. Risking this type of error is undesirable because it works in the direction of the hypothesis. However, by discarding these *Ss* the error risk is that of discarding some *Ss* who actually understood the instructions, but were just poor performers. Since there was no reason to suspect that poor performing retardates would behave more like normals on this task than the better performing retardates, it was felt that this selection on a performance basis would not work in favor of the hypothesized interaction.

RESULTS AND DISCUSSION

For each *S* the total number of *yes* responses (false positives) was de-

and the teachers at Dixon State School for their help and cooperation during this phase of the experiment.

²The author is grateful to R. Metzger

terminated for E words and C words. The means for retarded and normal Ss are presented in Figure 1. We may ask three questions of the data. First, considering only the college Ss, we may ask whether there were more false positives to E words than to C words. A paired-sample t test yields a value of 4.31 ($df = 29$, $p < .01$). Second, we may ask the same question of retardates, that is, did they also falsely recognize E words more often than C words? A t of 2.51 ($df = 38$, $p < .05$) indicates that they did. Thus, the results are consistent with the notion that an RR may be confused with an IAR. The third and most interesting question concerns the interaction. An analysis of variance indicated a significant interaction between Class of $S \times$ Type of Word ($F = 5.17$, $df = 1/69$, $p < .05$). Since there were several zero scores, the analysis of variance was repeated with each score

transformed by adding .5 to it and then taking its square root. The transformation left the basic findings unaltered. Finally, the analysis was repeated one additional time including the 10 retarded Ss which were discarded for failing to reach the 50% criterion. The inclusion of these Ss increased the magnitude of the interaction ($F = 9.62$, $df = 1/79$, $p < .01$).

The nature of the interaction may be seen in Figure 1. The absolute difference in mean number of false positives to E and C words for retardates appears considerably smaller than does the same comparison for college Ss. The interaction suggests that retardates may be responding with weaker or fewer IARs than normals, or that they show less RR-IAR confusion. The results are consistent with Wallace and Underwood (1964) and support the position that a characteristic of retarded Ss may be a defi-

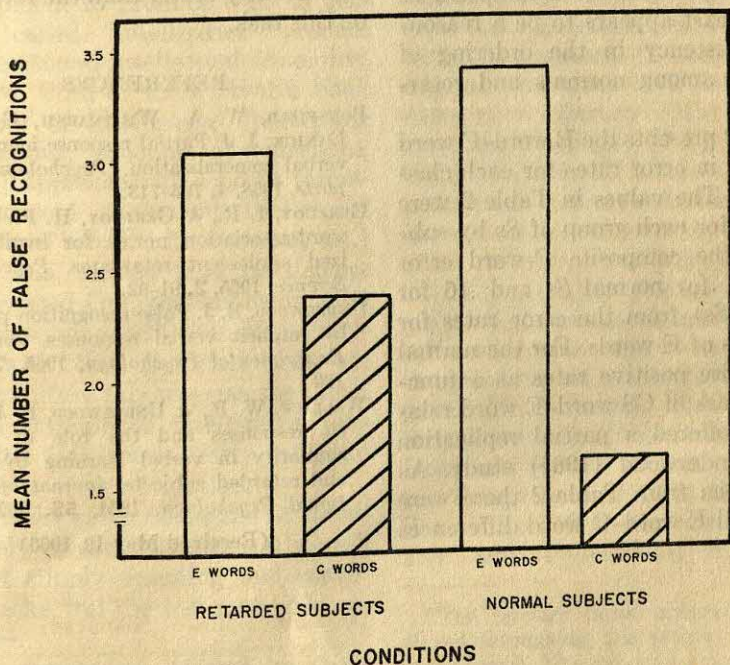


Fig. 1. Mean number of false positives to E words and C words for normal and retarded Ss.

TABLE 2
E WORD-C WORD DIFFERENCE IN ERROR
RATES AS A FUNCTION OF WORD CLASS

Subjects	Word class				
	A ₁	A ₂	CV	SO	SI
Normals	+ .01	+ .24	+ .19	+ .01	+ .05
Retardates	.00	+ .04	+ .07	+ .12	+ .02

ciency in making IARs. It must be recognized that other differences between normal and retarded Ss may be advanced to account for the results, but until a more compelling hypothesis is offered, it is suggested that differences in IAR processes were critical to the resultant interaction. It should be pointed out that the interpretation rests upon the assumption that there is a reasonable consistency between normals and retardates in their responses in word-association norms. A recent report by Gerjuoy and Gerjuoy (1965) renders this a feasible assumption as there at least appears to be a reasonable consistency in the ordering of responses among normals and retardates.

Table 2 presents the E word-C word difference in error rates for each class of words. The values in Table 2 were obtained for each group of Ss by subtracting the composite C-word error rates (.11 for normal Ss and .16 for retarded Ss) from the error rates for each class of E words. For the normal Ss the false positive rates as a function of class of CS word-E word relationship offered a partial replication of the Underwood (1965) study. As can be seen from Table 2 there were only small E word-C word differences

in error rates for Classes A₁ and SI, and there were substantial differences for Classes A₂ and CV. The one discrepancy from Underwood's data occurred in the failure of the present study to demonstrate increased false positives to E words of the SO class. The E word-C word differences in error rates for retardates were smaller than those for normals in all cases except Class SO. It is not clear why this one reversal occurred.

Each S made 100 responses and obtained a score of total correct recognitions (circling *yes* to a repeated item and circling *no* to a first occurrence). College Ss had a mean of 91.2 correct recognitions, and retarded Ss had a mean of 84.6 correct recognitions. The difference between means was significant ($t = 2.84$, $df = 69$, $p < .01$). Even though S-selection biases served to remove the poorest performing retardates, the normal Ss still performed significantly better than the retardates on this task.

REFERENCES

- BOUSFIELD, W. A., WHITMARSH, G. A., & DANICK, J. J. Partial response identities in verbal generalization. *Psychological Reports*, 1958, **4**, 703-713.
- GERJUOY, I. R., & GERJUOY, H. Preliminary word-association norms for institutionalized adolescent retardates. *Psychonomic Science*, 1965, **2**, 91-92.
- UNDERWOOD, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, **70**, 122-129.
- WALLACE, W. P., & UNDERWOOD, B. J. Implicit responses and the role of intralist similarity in verbal learning by normal and retarded subjects. *Journal of Educational Psychology*, 1964, **55**, 362-370.

(Received May 19, 1966)

SECOND-ORDER ABILITY STRUCTURE REVEALED IN RIGHTS AND WRONGS SCORES¹

JOHN L. HORN AND WILLIAM J. BRAMBLE

University of Denver

Using 14 ability tests to measure 12 primary factors in a sample of 106 adult males, 2nd-order factorial structures were determined separately for both rights scores and wrongs scores. In both analyses the results agreed well with previous analyses in showing intellectual functions interpreted as fluid intelligence, crystallized intelligence, general visualization, and general fluency. Other factors were interpreted tentatively. The correlations between rights scores and wrongs scores for the same test were found to range from +.51 to -.98, with most being in a range indicating about 25% of variance in common for the 2 kinds of scores. The results were interpreted as indicating that whereas the same basic functions were evidenced by the interrelationships among the 2 kinds of scores, decisions about the level of function of a particular individual could be quite disparate for rights as compared with wrongs scores.

In recent studies by Cattell (1963), Horn (1965) and Horn and Cattell (1966a; 1966b; 1967) evidence was found to support a theory of fluid and crystallized intelligence. According to this theory the influences operating in development to produce what is generally called "intelligence" produce not one general intellectual dimension, but two. Both of these involve basic processes of intelligence (perception of relations, education of correlates, span of apprehension, concept formation, concept attainment, the use of generalized solution instruments, etc.) but in one, termed fluid intelligence (abbreviated Gf), the emphasis is on problem solving in the immediate situation with test materials that are largely culture fair whereas the other, labeled crystallized intelligence (abbreviated Gc), more nearly indicates the limits of acculturation.

The tasks which define the Gf factor definitely require "intellectual work," are not simply scanning and movement tasks, but the test materials are

such that for most of the people tested, the fundamentals (i.e., the contents between which relations can be said to exist) are about equally common or equally novel. In terms of the primary abilities analyzed in the Horn and Cattell studies with adults, the processes and materials are best represented by Induction (I-Letter Series²), Figural Relations (CFR-Matrices), Associative Memory (Ma) and Figural Classifications (CFR). Similarly, "intellectual work" is indicated by the tasks which define Gc, but in this case the work involved is not so much that of the immediate problem-solving task as it is that which would have occurred previously through intensive acculturation to which only some, not all, of the people tested would have been exposed. The primaries which particularly characterized this function in the Horn and Cattell studies were Verbal Comprehension (V), Mechanical Knowledge (Mk), Numerical Facility (N) and Experiential Evaluation

¹This research was financed by Grant Number Ns G-518-618 from the National Aeronautics and Space Administration.

²The primary factor abbreviations employed throughout this article were taken from French, Ekstrom, and Price (1963), if they list the factor, or otherwise from French (1951) or Guilford and Merrifield (1960).

(EMS). The fluid and crystallized dimensions were found to be highly co-operative in General Reasoning (R), Semantic Relations (CMR-Word Analogies) and Formal Reasoning (Rs-Syllogisms).

Both of the above factors were clearly distinguished from a general visualization function (abbreviated Gv) and a general fluency dimension (abbreviated F). Gv involved all tasks which required work with spatial materials. It included the I and CFR primaries of Gf, but was defined principally by Spatial Orientation (S), Visualization (Vz), Speed of Closure (Cs), Flexibility of Closure (Cf), and Perceptual Speed (P). The general fluency function was manifested in all tasks requiring scanning or production of conventional concept labels (words), as best represented by Associational Fluency (Fa), Ideational Fluency (Fi), and Word Fluency (Fw).

Horn (1965) and Horn and Cattell (1966b, 1967) reviewed cross-sectional studies and produced new data to show that when putative measures of intelligence are classified as "primarily fluid," "primarily crystallized," and "about evenly mixed with these two," much of the seemingly contradictory evidence on age differences in intelligence can be seen to be consistent. Younger adults were found to be superior in fluid functions; older adults were found to be superior in crystallized functions and no systematic age differences were found for primaries and omnibus tests which involved these two factors in about the same degree.

This line of research thus shows promise of integrating previously diverse bits of evidence in the field of human abilities. Perhaps more interesting, it shows some promise of furthering rapprochement among previously separate, if not antagonistic,

subfields in which the principal concern is with questions about the processes of problem solving, perception, etc. But this promise can be realized only if the results obtained heretofore are not specific to either a particular sample of primary factors or a particular sample of subjects. It is of considerable interest, therefore, to test the generality of the above results. The main purpose of the present study is to do this.

Several years ago Fruchter (1950, 1953) effectively demonstrated that wrongs scores obtained from time-limit tests can have reliabilities and validities comparable to those for rights scores obtained from the same tests. More important, he showed that these results exist even when the correlations between rights and wrongs scores are quite low—some hovering near zero! In a factor analysis which included some rights-score and some wrongs-score variables he found a factor defined exclusively by wrongs scores. Horn (1965) replicated this finding and in his further analyses found that the carefulness factor (as Fruchter had labeled it) was largely independent of other factors even at the second and higher orders.

These findings thus suggest that quite different functions are measured with wrongs scores as compared with rights scores from timed ability tests.

But there are at least two ways in which to interpret this last statement. According to one interpretation it is evident from the fact of low correlation between comparable rights and wrongs scores that different functions are represented in these variables. For a low correlation means that a person who would be judged "good" in his performance on a given test when the rights score is used might be—that is, is nearly as likely to be—judged "poor" in his performance on this same test when the wrongs score is used.

Thus if "function" is thought of in terms of level of performance, it is indeed evident that different functions are tapped by rights and wrongs scores.

However, one might think of "function" not simply in terms of level of performance, but in terms of patterns of correlations of variables with all other variables with which they might be compared. It is possible for such patterns to be very similar for rights and wrongs scores even when the two kinds of scores correlate rather lowly. Horn and Cattell (1965) recently discussed this possibility under the heading of vehicles in measurement. A vehicle is an individual differences characteristic through which other characteristics are expressed. Thus carefulness may be systematically expressed in all of several kinds of abilities, but it need not be an influence which prevents the appearance of the factors representing these abilities when a properly rotated factor-analytic solution is obtained. In other words, if carefulness is a unitary function which affects all abilities in approximately the same way relative to a rights-score function, then it is possible for the same factors to appear in analyses of wrongs scores alone and in analyses of rights scores alone.

With respect to these considerations, then, the present investigation was designed to: (a) replicate previous findings with a new sample of subjects and a somewhat different sample of variables, and (b) see if the basic structure is found both among rights-score and among wrongs-score variables.

PROCEDURE

Selection of Variables

As in the previous studies of the fluid-crystallized concept, the French, Ekstrom, and Price (1963), French (1951), and Guilford and Merrifield (1960) collations of rotated factor results provided the basis for

selection of primary factor variables. To ensure some continuity with previous studies, nine primary factors used in the Horn and Cattell (1966a) analyses were selected for the present investigation. These are the variables identified with asterisks in Table 1, which provides basic descriptive data on all variables. One test was used to measure each primary factor. For P, Vz, Cf, and I the tests were entirely different from those used in the Horn-Cattell studies, although they represented the same primaries. The other tests were different only in the sense that they involved somewhat different items.

Three new primary factors and two new tests, for which the factor structure was not known, were added to the core variables described above. The Aiming primary, (A), was included on the hypothesis that it would help define general visualization, while two memory primaries—Memory Span (Ms) and Memory for Designs (Md)—were included on hypotheses that a major portion of their variances would go to fluid intelligence. The Dominoes Test, originated by Anstey and developed by Vautrain (1954), was included on the assumption that it would have both general visualization and fluid intelligence variance, but mainly the latter. S. Mednick's (1962)³ Remote Associations Test (RAT) was included as a measure of creativity. Burt (1962) has pointed out that useful creativity (in contrast to the originality of the insane, the delirious, the day-dreamer, etc.) involves the noegenetic processes of intelligence, as most fully described in Spearman's work. On this basis it was predicted that to the extent that the RAT indeed measures creativity, it should have some variance on the fluid-intelligence factor. Intuitively it would seem to involve some variance on the general-fluency factor as well.

Subjects and Administration

The subjects were 106 male inmates at the Colorado State Penitentiary.⁴ The tests were administered in 10 parts on 10 occa-

³The authors are very grateful to S. A. Mednick for making the materials for this test available to us prior to the publication, by Science Research Associates, of a commercial version of the RAT.

⁴The authors thank George Levy, Senior Psychologist at the Colorado State Penitentiary, for his help in securing this sample. In this respect they are also grateful to Wayne Patterson, Warden, and Harry Tinsley, Director of Institutions.

TABLE 1
PRIMARY FACTOR VARIABLES, TESTS USED TO ESTIMATE THEM, AND MEANS
AND STANDARD DEVIATIONS

Primary factor	Symbol	Test used	M	Sigma
1. Perceptual Speed	*P	Cancel Numbers	515.9	73.00
2. Aiming	A	Dot in Circle	1145	386.0
3. Visualization	*Vz	Follow the Line	425.6	102.8
4. Flexibility of Closure	*Cf	Copy the Pattern	110.5	50.41
5. Figural Relations	*CFR	Figure Series	37.49	7.429
6. Deduction	D	Dominoes	28.76	8.086
7. Memory Span	Ms	Forward-Backward Digits	434.6	188.6
8. Memory for Designs	Md	Redraw the Figure	4.868	5.206
9. Induction	*I	Letter Series	58.06	15.82
10. Semantic Relations	*CMR	Common Word Analogies	22.06	5.449
11. Creativity	RAT	Remote Associations	86.79	32.75
12. Associational Fluency	*Fa	Similar Words	66.74	29.63
13. Ideational Fluency	*Fi	Things	103.9	40.78
14. Verbal Comprehension	*V	General Information	36.12	7.683
Word Variables				
15. Perceptual Speed	P	Cancel Numbers	21.67	16.48
16. Aiming	A	Dot in Circle	413.5	334.4
17. Visualization	Vz	Follow the Line	9.934	33.54
18. Flexibility of Closure	Cf	Copy the Pattern	56.32	43.84
19. Figural Relations	CFR	Figure Series	22.54	7.417
20. Deduction	D	Dominoes	24.09	8.238
21. Memory Span	Ms	Forward-Backward Digits	45.36	43.89
22. Induction	I	Letter Series	15.11	10.50
23. Semantic Relations	CMR	Common Word Analogies	18.38	5.141
24. Creativity	RAT	Remote Associations	19.65	14.67
25. Associational Fluency	Fa	Controlled Associations	41.65	22.81
26. Ideational Fluency	Fi	Things	5.207	6.334
27. Verbal Comprehension	V	General Information	22.31	6.709

sions spaced over 5 days. There was one testing session in the morning, one in the afternoon on each day. A testing session lasted about 1½ hours.⁶ The men were volunteers. They were paid \$2.00 each for their services.⁶ The morale and motivation was judged to be very high throughout, although somewhat lower in the later sessions than in the earlier ones.

ANALYSES AND RESULTS

The number of correct answers and the number of incorrect answers was determined for each part-test. A vari-

able was then obtained as a simple (unweighted) sum of the raw scores on the 10-part tests. All variables, both rights and wrongs scores, were inter-correlated by the product-moment formula. The resulting table of inter-correlations has been deposited with ADI.⁷ In Table 2 are shown the correlations between rights and wrongs scores for the same test.

The submatrices involving rights

⁶ Marlan Wilson, Jr. directed the administration and scoring of the tests. He was assisted by Messrs. Mathis and Conrad. The help of these men is very gratefully acknowledged.

⁶ We were told that on the prison-to-street exchange, this is equivalent to about \$10.00 to \$15.00 outside the prison.

⁷ Material supplementary to this article has been deposited with the American Documentation Institute. Order Document No. 9274 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 2
CORRELATIONS BETWEEN RIGHTS AND WRONGS SCORES FOR THE SAME TEST

	Variable symbol												
	P	A	Vz	Cf	CFR	D	Ms	I	CMR	RAT	Fa	Fi	V
<i>r</i>	.12	-.54	-.52	-.55	-.98	-.63	.30	-.64	-.98	-.40	.51	.46	-.92

scores alone and wrongs scores alone were abstracted and factored separately. Algebraic independence was maintained among the variables of these two sets. In each set the squared multiple correlation of a variable with the other variables of the set was determined and inserted in the principal diagonal as a communality estimate. Principal axes factors were then extracted until a latent root became zero. The number of factors to be rotated was further limited by selecting only those factors with roots larger than one which had at least one loading greater than .20 (absolute value)—that is, taking factors in order of size of latent roots, when a principal axis factor was encountered which did not

have at least one loading greater than .20 (absolute value), that factor and all factors with smaller latent roots were eliminated in the rotations. The factors were rotated in accordance with the criteria of Kaiser's (1958) varimax procedure.

The rotated results from the factor analyses are shown in Tables 3 and 4. The squared multiple correlation communalities are shown in the next-to-last column in each of these tables; the communalities based on factor loadings are shown at the far right.

DISCUSSION

The correlations of Table 2 suggest that when a task is highly speeded but conceptually simple, (Canceling,

TABLE 3
ROTATED FACTORS FROM ANALYSIS OF RIGHTS SCORES

Symbols	Primary factors	Second order factors							
		I	II	III	IV	V	VI	SMR	λ^2
P	Circle Numbers	11	01	59	12	09	—06	31	39
A	Aiming	18	11	50	18	16	41	45	52
Vz	Visualization	25	07	67	08	18	23	52	62
Cf	Flexibility of Closure	45	28	37	24	32	32	63	67
CFR	Figural Series	61	03	31	10	35	18	59	63
D	Deduction	69	38	25	10	13	04	65	71
Ms	Memory Span	12	04	12	11	49	01	23	28
Md	Memory Designs	22	09	18	11	53	39	45	53
I	Letter Series	57	42	22	27	40	09	74	79
CMR	Common Word Analogies	39	51	16	16	48	06	67	73
RAT	Creativity	25	61	—01	15	—06	10	42	48
Fa	Controlled Associations	16	41	15	70	35	—04	76	84
Fi	Things	13	28	24	74	09	18	69	74
V	General Information	05	75	07	28	18	01	63	69

Note.—Decimal points have been omitted in the factor coefficients in order to simplify the presentation.

TABLE 4
ROTATED FACTORS FROM ANALYSIS OF WRONGS SCORES

Symbols	Primary factors	Second order factors						
		I	II	III	IV	V	SMR	h^2
P	Circle Numbers	10	11	58	—06	39	45	52
A	Aiming	16	03	56	—02	—04	32	35
Vz	Visualization	14	04	65	—03	07	37	45
Cf	Flexibility of Closure	46	33	28	—04	43	55	59
CFR	Figural Series	55	—11	31	—18	25	45	51
D	Deduction	69	31	15	13	03	60	62
Ms	Memory Span	25	17	—11	22	03	15	15
I	Letter Series	73	22	32	05	—11	63	69
CMR	Common Word Analogies	61	49	19	—15	24	66	73
RAT	Creativity	14	32	—15	61	—26	47	58
Fa	Controlled Associations	07	—09	—05	72	07	45	54
Fi	Things	—13	—11	06	61	—04	33	40
V	General Information	26	70	10	—06	04	50	57

Note.—Decimal points have been omitted in the factor coefficients in order to simplify the presentation.

for example), the number of errors correlates positively with the number of correct responses.

In the Canceling test the wrongs score is literally an error score, but this is not quite true for the other three tests for which there is positive correlation between rights and wrongs scores. In Memory for Digits the subject was required to recall the digits in the order in which they were given (in half the items) or in reverse order. If any one of the digits given in an answer was incorrect or out of order, the response on that item was scored wrong. In fluency tests (Similar Words and Things) a response was judged "wrong" if it seemed to be a quite bizarre, irrelevant association. For example, the response "clip" to the key word "warm" was scored wrong. In some cases, of course, one could legitimately question the decision that a particular response was bizarre or irrelevant. In this connection it is perhaps worth pointing out that the correlations between RAT (a measure of creativity) and the wrongs scores for Fa and Fi were $-.02$ and $-.06$ re-

spectively—that is, certainly low enough to suggest that relevant but remote associations were not being counted in the wrongs scores for Fa and Fi.

The correlations between wrongs and rights scores are negative and very high for Figure Series, Common Word Analogies, and General Information. This indicates that these measurements were obtained under almost power conditions, as was the aim in this research. Had all subjects attempted all items in these tests, the correlations between rights and wrongs scores for the same test would necessarily be -1.0 . The departure of the correlations from -1.0 is thus a rough indication of the speededness of test administration.

The correlations between rights and wrongs scores for A, Vz, Cf, D, I, RAT, Fa, and Fi indicate that the two kinds of scores have about 25% of their variance in common. This is certainly low enough to suggest that rights and wrongs scores represent somewhat distinct functions. It is therefore particularly interesting that

the factorial results for the analyses of these two kinds of scores are so similar. Although there are differences in detail, the same essential functions, replicating previous results, can be identified in both analyses. The first four factors are easily identified as the fluid intelligence, crystallized intelligence, general visualization, and general fluency dimensions previously identified by Horn and Cattell (1966a).

The fifth factor in the rights score analysis is characterized mainly by the memory tests but involves substantial loadings on CMR, I, CFR, Fa, and Cf. Each of these tests requires that the subject hold a relation or several relations in immediate awareness in order to compare it with other relations. This suggests a link between this factor and the process which Spearman first described as span of apprehension or what has been discussed in recent years under the heading of temporal integration.

Factor VI in the rights score analysis represents a visualization function but in this case, in contrast to Factor III, Memory for Designs is involved. There is thus a suggestion that visual retention, as represented in Factor VI, is somewhat distinct from the visual "fluency" represented in Factor III.

The simple structure in the analysis of wrongs scores is somewhat better than in the rights score solution: yet the salient factor coefficients tend to be of about the same magnitude. The suggestion is that analysis of wrongs scores yields a "cleaner" structure than analysis of rights scores. It is of some interest to compare the differences in loadings for the comparable salient variables in the rights-score and wrongs-score analyses, but it is probably not worthwhile to occupy space with a detailed discussion of these at this time.

It is perhaps worth noting, however, that Remote Associations in the rights-score analysis has a loading of only .15 on the general-fluency dimension, whereas in the wrongs-score analysis the comparable loading is .61. General fluency is interpreted as a facility in bringing concept labels (words) from a long-term storage center into immediate awareness. The results obtained here suggest that this facility is not characterized by criticality. That is, the concepts implied need not be related in a precise and logical manner, for this is the requirement imposed on associations in the rights score for RAT and this variable does not help to define F, whereas this restriction is removed in the wrongs score for RAT and in this case the variable does help to define F.

Factor V in the wrongs-score analysis does not match any factor found in the rights-score analysis or in our previous work. It is not readily interpreted. We will therefore hold our attempt at identifying it until after we find that it can be replicated.

Something of a paradox is thus posed by these results. When a test is scored by the wrongs-score procedure, the person who adopts a strategy of avoiding errors has the advantage; when the same test is scored by the rights-score procedure, the person who adopts a strategy of getting as many right as possible, even at the cost of a few errors, has the advantage. The same basic intellectual functions are indicated by analyses of both kinds of scores. This means that regardless of strategy adopted, knowing how a subject performs on one kind of task in a factor pattern, we can predict with considerably better than chance accuracy how he will perform in all other tasks defining the pattern. Yet, and here is the paradox, persons who are rated high in a particular func-

tion when this is defined in terms of rights scores will rather frequently (as suggested by correlations of the order of .5) be rated comparatively low in this same function when it is defined in terms of wrongs scores (that is, a constant minus number of wrongs, to define the variable so that a high score indicates good performance).

This is indeed an interesting finding and one which has practical as well as theoretical implications. It is to be hoped that follow-up studies now in progress will shed additional light on the questions here provoked.⁷

REFERENCES

- CATTELL, R. B. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 1963, **54**, 1-22.
- FRENCH, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monographs*, 1951.
- FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. *Manual for kit of reference tests for cognitive factors*. Princeton, New Jersey: Educational Testing Service, 1963.
- FRUCHTER, B. Error scores as a measure of carefulness. *Journal of Educational Psychology*, 1950, **41**, 279-291.
- FRUCHTER, B. Differences in factor content of rights and wrongs scores. *Psychometrika*, 1953, **18**, 257-265.
- GUILFORD, J. P., & MERRIFIELD, P. R. The structure of intellect model: Its uses and implications. Reports from the Psychological Laboratory, University of Southern California, 1960.
- HORN, J. L. Fluid and crystallized intelligence: A factor analytic and developmental study of structure among primary mental abilities. Unpublished doctoral dissertation, University of Illinois, 1965.
- HORN, J. L. & CATTELL, R. B. Age differences in primary mental abilities. *Journal of Gerontology*, 1966, 210-220. (a)
- HORN, J. L., & CATTELL, R. B. Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 1966, **57**, 253-270. (b)
- HORN, J. L. & CATTELL, R. B. Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 1967 (in press).
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187-200.
- MEDNICK, S. A. The associative basis of the creative process. *Psychological Review*, 1962, **69**, 220-232.
- VAUTRIN, H. *The D48 test*. Revised and adapted from the French by John D. Black. Palo Alto, California: Consulting Psychologists Press, 1954.

(Received May 23, 1966)

ERRATUM

On page 332 of "Structure of Intelligence in Negro and White Children" by Ira J. Semler and Ira Iscoe in the December 1966 issue the last sentence in the section WISC Structure: Multivariate Analysis of Variance and Exploratory Factor Analyses should read: Factor II for the white sample was a complex factor involving loadings of the WISC A, DS, BD, and C subtests.

DISCUSSION VERSUS MEMORY IN COOPERATIVE GROUP CONCEPT ATTAINMENT

PATRICK R. LAUGHLIN AND MARY AUSTIN DOHERTY

Loyola University

In order to determine the relative importance of discussion vs. memory in group-concept attainment, the performance of 2-person cooperative female groups was investigated in 5 concept-attainment problems. A $2 \times 2 \times 2 \times 2$ factorial design was used with the variables: (a) discussion (allowed or not allowed), (b) memory (paper allowed or not allowed), (c) stimulus display (form or sequence), (d) number of relevant problem attributes (2 or 4). Discussion resulted in fewer card choices to solution, lower percentage of untenable hypotheses, and more time to solution, while memory had no effect. There were no main effects for the use of focusing or scanning strategies, although both strategies showed similar complex relationships in the interactions of the 4 variables.

In a study of concept attainment by individuals versus two-person groups, Laughlin (1965) found that groups required fewer card choices to solution and used the focusing strategy more than individuals. Although this group superiority in card choices was not found using the Taylor-McNemar (1955) correction model, the greater use of focusing by groups remained even after the correction. Focusing strategy is distinguished from scanning as one of two basic problem-solving processes adopted by subjects in a concept-attainment situation by Bruner, Goodnow, and Austin (1956). In focusing the subject tests the relevance of all the possible hypotheses involved in a particular attribute or attributes by choosing a card differing in one (conservative focusing) or more (focus gambling) attributes from a positive focus card. In scanning he tests specific hypotheses, either singly (successive scanning), all at once (simultaneous scanning), or some intermediate number. In general, focusing is a more effective strategy, which Bruner et al. (1956) interpret as due to the greater inference and memory requirements of scanning.

Since memory and inference requirements seem to be two basic elements

of a concept-attainment task, group superiority may be due to either better memory, better inference, or both. The important effects of memory in concept learning have been demonstrated in experiments by Hovland and Weiss (1953), Cahill and Hovland (1960), Hunt (1961), Bourne, Goldstein, and Link (1964), and extensively reviewed by Dominowski (1965). Thus, the superiority of groups may have been due to their better memory. An alternative explanation involves the inference or logic requirements of the concept-attainment problem, in which the discussion process between the individuals may have enabled them to evaluate the information on each card choice, reason concerning the meaning of positive and negative instances, eliminate errors in inference by monitoring each other's card choices and hypotheses, and so forth. These processes can be summarized as the effects of discussion as opposed to memory.

Thus, in order to test these two alternative explanations of memory and discussion for the superiority of groups over individuals, the present study compared two-person cooperative groups in which paper (memory) and discussion were or were not allowed in

a factorial design. In order to extend and replicate previous research (Laughlin, 1965, 1966) two additional variables were included: (a) form versus sequence stimulus displays, (b) two-attribute versus four-attribute problems. Finally, the subjects were females rather than the males of previous research.

METHOD

Design

A $2 \times 2 \times 2 \times 2$ factorial design was used with the following variables: (a) discussion (allowed or not allowed), (b) memory (paper allowed or not allowed), (c) stimulus display (form or sequence), (d) number of relevant attributes (two to four).

Subjects

The subjects were 192 female college students at Alverno College, Milwaukee. Six pairs were randomly assigned to each of the 16 experimental conditions.

Stimulus Displays and Problems

The problem materials were six 28-inch $\times$ 44-inch white poster boards, each containing an 8×8 array of 64 $2\frac{1}{2}$ -inch by 4-inch cards drawn in colored ink with dark outlines. The 64 cards represented all the possible combinations of six attributes with two levels of each attribute. Form displays consisted of the following attributes and values: (a) shape: square or triangle, (b) size: large or small, (c) number: one or two, (d) color: red or green, (e) pattern: striped or solid, (f) borders: one or two. Sequence displays consisted of all combinations of six plus and/or minus signs in a row. In order to facilitate reference to the six positions, each was a different color, so that the color name was the attribute and plus or minus the value of each color. The attributes and values were listed on reference cards which the subjects could use throughout the experiment. Three different ordered displays, in which the position of each card varied systematically in relation to the other cards (e.g., all red figures on the top four rows), were used for both form and sequence stimulus displays. Each problem and initial card for each pair of subjects were randomly selected from the total subsets of possible two-attribute or

four-attribute concepts. All pairs solved five problems.

Procedure

The instructions demonstrated the meaning of conjunctive concepts, pointed out the attributes and values on the stimulus display, explained that the problems all involved two (or four) relevant attributes, and emphasized that they were to be solved in as few card choices as possible, regardless of time (Laughlin, 1964). It was emphasized in all conditions that the two persons were not competing in any way, but would be scored as a cooperative group and compared with other groups. In the no-discussion conditions either of the two persons could make each card choice and hypothesis, and the two persons were seated adjacently so that they saw and heard both each other and the experimenter, but could not speak with each other. They were instructed to listen carefully to each other's card choices and hypotheses, and to attempt to utilize this information.

RESULTS

The means for the 16 treatment groups for (a) number of card choices to solution, (b) percentage repetitions of hypotheses, (c) percentage of untenable hypotheses, (d) time to solution, (e) focusing strategy, and (f) scanning strategy are given in Table 1. All measures are for totals over the five problems, and totals over both subjects regardless of which subject made the card choice or hypothesis.

Card Choices to Solution

Two-person groups who were allowed discussion had fewer card choices to solution than those not allowed discussion, $F(1, 80) = 6.91$, $p < .02$. There was no difference between those who used paper and those who did not, $F < 1$. Likewise, the effects of stimulus display and number of problem attributes were not significant, $F(1, 80) < 1$; $F(1, 80) = 2.54$, respectively. The Stimulus Display $\times$ Attributes interaction was significant, $F(1, 80) = 4.02$, $p < .05$,

TABLE 1
MEAN NUMBER OF CARD CHOICES TO SOLUTION, PERCENTAGE REPETITION OF
HYPOTHESES, PERCENTAGE UNTENABLE HYPOTHESES, TIME TO SOLUTION,
FOCUSING STRATEGY, AND SCANNING STRATEGY, OVER FIVE PROBLEMS

	Paper				No paper			
	Form		Sequence		Form		Sequence	
	Two	Four	Two	Four	Two	Four	Two	Four
Discussion								
Card choices	18.33	22.50	20.67	21.67	28.50	22.67	15.83	24.67
Repetition hypotheses	.05	.08	.01	.08	.08	.07	.04	.12
Untenable hypotheses	.17	.39	.16	.21	.31	.26	.20	.30
Time to solution	20.8	31.3	16.7	22.8	20.2	43.3	19.8	31.5
Focusing strategy	3.56	2.42	2.45	2.30	2.31	3.21	3.25	2.55
Scanning strategy	21.76	22.32	16.94	19.06	16.12	18.79	25.01	17.13
No discussion								
Card choices	30.33	25.83	23.50	28.67	21.50	25.33	22.00	27.33
Repetition hypotheses	.07	.06	.06	.12	.07	.19	.22	.09
Untenable hypotheses	.44	.39	.23	.31	.24	.48	.28	.49
Time to solution	15.5	27.7	13.2	26.2	14.0	23.3	13.2	20.5
Focusing strategy	2.25	2.61	2.72	2.61	3.11	2.39	2.48	1.82
Scanning strategy	16.72	16.78	18.30	18.96	18.96	18.51	18.01	16.77

Note.—Time in minutes. Maximum focusing strategy is 5.00.

so that two-attribute and four-attribute form problems showed little difference, while two-attribute sequence problems were solved in fewer card choices than four-attribute sequence problems.

Percentage Repetitions of Hypotheses

In order to control for different numbers of hypotheses across conditions, the percentage of repetitions of hypotheses was determined by dividing the number of repetitions of hypotheses by the number of hypotheses. Both the effects of discussion and memory were significant. Discussion groups had a lower percentage of repetitions of hypotheses than no-discussion groups, $F(1, 80) = 4.13, p < .05$; and paper groups had a lower percentage than no-paper groups, $F(1, 80) = 3.97, p < .05$. The effects of stimulus display and number of problem attributes were not significant.

Percentage Untenable Hypotheses

Untenable hypotheses were of two types: (a) a hypothesis for a value of an attribute when the other value of the attribute had previously occurred on a positive instance, for example, the hypothesis "red square" when a green instance had been positive; (b) a hypothesis which had previously occurred on a negative instance, for example, the hypothesis "red square" when an instance with a red square had been negative. In order to control for the different number of hypotheses across conditions, the number of untenable hypotheses was divided by the number of hypotheses. Discussion groups had a lower percentage of untenable hypotheses than no-discussion groups, $F(1, 80) = 12.32, p < .001$; paper and no-paper groups did not differ, $F(1, 80) = 1.05$; form displays had a lower percentage than sequence displays, $F(1, 80) = 4.43, p < .05$; and two-attribute problems

had a lower percentage than four-attribute problems, $F(1, 80) = 10.60$, $p < .001$.

Time to Solution

Discussion groups required more time to solution than no-discussion groups, $F(1, 80) = 7.13$, $p < .01$; there was no difference between paper and no-paper groups, $F(1, 80) < 1$; there was no difference between form and sequence displays, $F(1, 80) = 2.74$; and four-attribute problems required more time than two-attribute, $F(1, 80) = 22.57$, $p < .001$.

Focusing Strategy

Focusing strategy was scored according to two rules. Rule 1: Each card choice had to obtain information on one new attribute. New information was obtained if the card choice altered only one attribute not previously proven irrelevant (conservative focusing), or, if more than one attribute was altered (focus gambling), the instance was either positive or the ambiguous information correctly resolved on the next card by altering only one attribute. Rule 2: If a hypothesis was made it had to be tenable considering the information available. Each card choice and accompanying hypothesis that satisfied these two rules was counted as an instance of focusing, and the total number of such instances was divided by the total number of card choices on the problem to give a continuous focusing score from .00 to 1.00.

None of the four main effects of discussion, memory, stimulus display, or attributes were significant, $F(1, 80) = 2.18$; $F < 1$; $F < 1$; $F = 2.52$, respectively. The Discussion $\times$ Memory $\times$ Stimulus Display interaction was significant at the .05 level, $F(1, 80) = 5.18$. With discussion and paper there was more focusing with form

displays than with sequence displays, while with discussion and without paper there was no difference between form and sequence displays. Without discussion the relationships were reversed, so that with paper there was no difference between form displays and sequence displays, while without paper there was more focusing with form displays. The Discussion $\times$ Memory $\times$ Attributes interaction was also significant at the .05 level, $F(1, 80) = 4.97$, and showed the same relationships. With discussion and paper there was more focusing with two-attribute problems than with four-attribute, while without paper there was no difference between two-attribute and four-attribute problems. Without discussion the relationships were reversed, so that with paper there was no difference between two-attribute and four-attribute problems, while without paper there was more focusing with two-attribute problems. Finally, the four-way interaction was significant, $F(1, 80) = 4.88$, $p < .05$.

Scanning Strategy

Scanning strategy was scored by comparing each card in turn with the given problem card. If the selected card was positive, all concepts different on the given and selected cards were eliminated; if the selected card was negative, all concepts identical on the given and selected cards were eliminated. The total of the number of concepts thus eliminated plus those concepts eliminated by direct hypotheses was then divided by the total number of card choices on the problem in order to determine the average number of concepts eliminated per card choice. This measure was considered an index of scanning.

There was only a slight trend for discussion groups to have higher scanning scores at the .10 level, $F(1, 80) =$

TABLE 2
INTERCORRELATIONS OF RESPONSE
MEASURES

	Card choices	Repe- tition hy- poth- eses	Un- tena- ble hy- poth- eses	Time to so- lution	Focus- ing strategy
Repetition hypotheses	.30				
Untenable hypotheses	.52	.24			
Time to so- lution	-.33	.12	.25		
Focusing strategy	-.54	-.12	-.23	-.02	
Scanning strategy	-.68	-.14	-.14	-.14	.51

2.68. The effects of memory, stimulus display, and attributes were all non-significant, all F s < 1. The Discussion $\times$ Memory $\times$ Stimulus Display interaction was significant, $F(1, 80) = 6.38$, $p < .02$. The relationships were parallel to those for focusing but more pronounced. With discussion and paper there was more scanning with form displays than with sequence displays, while without paper there was more scanning with sequence displays. Without discussion the relationships were reversed, so that with paper there was more scanning with sequence displays, while without paper there was more scanning with form displays.

Correlations between Response Measures

Correlations between the six response measures are given in Table 2.

DISCUSSION

The major purpose of the experiment was to assess the relative importance of discussion and memory as potential factors in the previously demonstrated superiority of group over individual concept attainment.

In terms of the basic dependent variable, number of card choices to solution as an indicator of efficiency of performance, discussion was clearly the more important factor, as discussion groups had fewer card choices than no-discussion groups while there was no difference for paper and no-paper groups. The same relationship was found for percentage of untenable hypotheses. Basically, the benefits of discussion seem to involve an inference and monitoring process, in which the two persons can both reason concerning the meaning of each card choice and hypothesis and check each other. Through discussion they can reduce erroneous inference and insure more efficient card choices, thus solving the problems in fewer card choices and making fewer untenable hypotheses. However, discussion groups required more time than no-discussion groups, while there was no difference between paper and no-paper groups. This result parallels the greater time required by groups than by individuals (Laughlin, 1965).

Discussion and memory did not result in significantly different use of either focusing or scanning strategies, although there were trends favoring discussion groups over no-discussion groups on both strategies. The differences in card choices and untenable hypotheses as a function of discussion, the lack of difference as a function of memory, the high correlation between focusing and scanning, and the parallel interactions between discussion, memory, and display for the two strategies seem to indicate the following overall conclusion: Discussion results in more effective problem solving rather than specific changes in the use of focusing or scanning, while memory has no effect. In other words, whether focusing or scanning was used, discussion enabled it to be used more

effectively while paper made no difference.

Contrary to previous research (Bourne, 1957, 1966; Bourne & Haygood, 1958; Laughlin, 1965; Shepard, Hovland, & Jenkins, 1961), there were no main effects indicating an increase in number of card choices with increased task complexity for four-attribute concepts versus two-attribute or sequence versus form displays. However, the Display $\times$ Attributes interaction was significant, so that there was no difference between two-attribute and four-attribute form problems, while four-attribute sequence problems required more card choices than two-attribute sequence problems. Thus, the combination of both the more complex sequence display and four-attribute problems resulted in an increased number of card choices although neither displays nor number of attributes alone resulted in a complexity effect. Since previous research has involved individuals working alone (except group conditions in Laughlin, 1965), the increased complexity necessary to demonstrate increased card choices may indicate a social facilitation effect from the mere presence of the other person, apart from the effects of discussion (Zajonc, 1965).

REFERENCES

- BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *Journal of Experimental Psychology*, 1957, **54**, 201-207.
- BOURNE, L. E., JR. *Human conceptual behavior*. Boston: Allyn and Bacon, 1966.
- BOURNE, L. E., JR., GOLDSTEIN, S., & LINK, W. E. Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology*, 1964, **67**, 439-448.
- BOURNE, L. E., JR., & HAYGOOD, R. C. The role of stimulus redundancy in the identification of concepts. *Journal of Experimental Psychology*, 1959, **58**, 232-238.
- BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
- CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology*, 1960, **59**, 137-144.
- DOMINOWSKI, R. Role of memory in concept learning. *Psychological Bulletin*, 1965, **63**, 271-280.
- HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology*, 1953, **45**, 175-182.
- HUNT, E. B. Memory effects in concept learning. *Journal of Experimental Psychology*, 1961, **62**, 598-604.
- LAUGHLIN, P. R. Speed versus minimum-choice instructions in concept attainment. *Journal of Experimental Psychology*, 1964, **67**, 596.
- LAUGHLIN, P. R. Selection strategies in concept attainment as a function of number of persons and stimulus display. *Journal of Experimental Psychology*, 1965, **70**, 323-327.
- LAUGHLIN, P. R. Selection strategies in concept attainment as a function of number of relevant problem attributes. *Journal of Experimental Psychology*, 1966, **71**, 773-776.
- SHEPARD, R. N., HOVLAND, C. I., & JENKINS, H. N. Learning and memorization of classifications. *Psychological Monographs*, 1961, **75**(13, Whole No. 517).
- TAYLOR, D. W., & MCNEMAR, O. W. Problem solving and thinking. *Annual Review of Psychology*, 1955, **6**, 455-482.
- ZAJONC, R. D. Social facilitation. *Science*, 1965, **149**, 269-274.

(Received June 9, 1966)

Introducing an Original Paperbound Series . . .

**CURRENT ISSUES AND RESEARCH
IN EDUCATION**

General Editor: Harry L. Miller, *Hunter College*

The *Current Issues and Research in Education* series is an innovation that, in the struggle to keep up with the "information explosion" in the rapidly expanding field of education, encompasses and illuminates a wide range of new ideas, experimentation, and criticism. Each volume contains excerpts, summaries, and entire articles that survey the most recent research findings, comment on persistent issues, and evaluate ongoing experiments and new ideas for the future. The editors' introductions provide the necessary historical background and underscore the relationships of the articles to general trends in the field. Throughout, special emphasis is placed on the significance of these issues for teachers in today's schools.

Volumes now available

**THE PSYCHOLOGY
OF EDUCATION**

**Edited by Donald H. Clark,
*Hunter College***

**EDUCATION FOR THE
DISADVANTAGED**

**Edited by Harry L. Miller,
*Hunter College***

**ELEMENTARY
EDUCATION**

**Edited by Maurie Hillson,
*Rutgers University***

**SOCIAL FOUNDATIONS
OF EDUCATION**

**Edited by Dorothy Westby-Gibson,
*San Francisco State College***

Just published

All volumes \$2.95 each

**BEHAVIOR IN INFANCY AND
EARLY CHILDHOOD**

A Book of Readings Edited by Yvonne Brackbill, *University of Denver*, and George G. Thompson, *Ohio State University*

Presenting the work of more than sixty of the world's leading authorities in the field, this comprehensive reader examines every aspect of development during infancy and early childhood—from the psychological to the social. Ranging from established classics to the first English translations of pioneering Russian studies, the selections present the full scope of contemporary research and are organized to encourage comparative examination.

Just published

704 pages, *illus.*

\$9.95

Available at your bookstore or directly from . . .

THE FREE PRESS

**A Division of The Macmillan Company
866 Third Avenue, New York 10022**

nine ways to enrich your understanding



EDUCATIONAL PSYCHOLOGY IN THE CLASSROOM

Third Edition

By HENRY CLAY LINDGREN, *San Francisco State College*. Focusing on the teacher as a behavioral scientist, the new edition of this successful book is concerned mainly with the ongoing classroom situation. New features include a unique chapter on the slum child; strong emphasis on the contributions of social psychology; a comprehensive treatment of field theory and its application to educational practices and problems; inclusion of such relatively-ignored topics as social learning theory, creativity, and needs for affiliation and achievement. 1967. 686 pages. \$8.50

INTRODUCTION TO STATISTICAL ANALYSIS AND INFERENCE

For Psychology and Education

By SIDNEY J. ARMORE, *George Washington University*. Provides a secure foundation in both descriptive statistics and statistical inference. All topics are supported by detailed explanations, illustrative problems, graphs and charts, and a review of algebraic symbols and operations is set forth in an appendix. 1966. 546 pages. \$8.95.

BEHAVIORAL SCIENCE FRONTIERS IN EDUCATION

Edited by ELI M. BOWER, *National Institute of Mental Health, Bethesda, Maryland*; and WILLIAM G. HOLLISTER, *University of North Carolina School of Medicine*. Prominent behavioral scientists—theorists and practitioners—set forth the germinal ideas and programs which they represent. 1967. Approx. 544 pages. Prob. \$8.95.

JOHN WILEY & SONS, Inc.
605 Third Avenue, New York, N.Y. 10016

of better teaching techniques



GROUPING IN EDUCATION

Edited by ALFRED YATES. Sponsored by the UNESCO Institute for Education, Hamburg, Germany. Describes and evaluates commonly-practiced grouping procedures, and presents accounts of research into grouping practices in the form of four special articles and fifty research abstracts. 1966. 314 pages. \$12.75.

CLASSROOM GROUPING FOR TEACHABILITY

By HERBERT A. THELEN, *University of Chicago*; with the assistance of HENRY PETERSON, ALAN OPPENHEIM, WILLIAM HOOCK, STEVEN PERLS, and ANNE BRODY. 1967. Approx. 288 pages. \$7.50.

READINGS IN THE PSYCHOLOGY OF PARENT-CHILD RELATIONS

Edited by GENE R. MEDINNUS, *San Jose State College*. 1966. 371 pages. \$4.50.

STUDIES IN COGNITIVE GROWTH

By JEROME S. BRUNER, *Harvard University*; with ROSE R. OLVER, *Amherst College*; PATRICIA M. GREENFIELD et al. 1966. 343 pages. \$7.95.

THE COMPUTER IN AMERICAN EDUCATION

By DON D. BUSHNELL, *Brooks Foundation, Santa Barbara*; and DWIGHT W. ALLEN, *Stanford University*. 1967. Approx. 256 pages. Cloth: Prob. \$6.95. Paper: Prob. \$3.95.

INTERNATIONAL STUDY OF ACHIEVEMENT IN MATHEMATICS

A Comparison of Twelve Countries

Volumes I and II

Chairman and Editor: TORSTEN HUSEN, *University of Stockholm*. 1967. Prob. \$19.95 the set. In press.

JOHN WILEY & SONS, Inc.

605 Third Avenue, New York, N.Y. 10016

Illiteracy Can Be Pretty Painful

But it doesn't have to be incurable.

Not when a special program like SRA's **READING IN HIGH GEAR** has helped functionally illiterate individuals to read at sixth-to-eighth grade levels with 150 to 300 hours of instruction.*

READING IN HIGH GEAR is a modern program particularly applicable for helping mature semi-literate and illiterate students learn the basic communicative skills which they need to make their lives more useful.

The stories used in the program are planned to be interesting to the student — relevant to his own experiences and his own age group. The story themes are such that they can help develop positive social attitudes and a healthy orientation to society.

READING IN HIGH GEAR is a versatile program, and can be effectively adapted for use with remedial students and pre-dropouts.

For many people, learning to read can be the all-important first step toward a better life. Write for a detailed booklet or ask your nearest SRA Staff Associate to tell you more about **READING IN HIGH GEAR**.

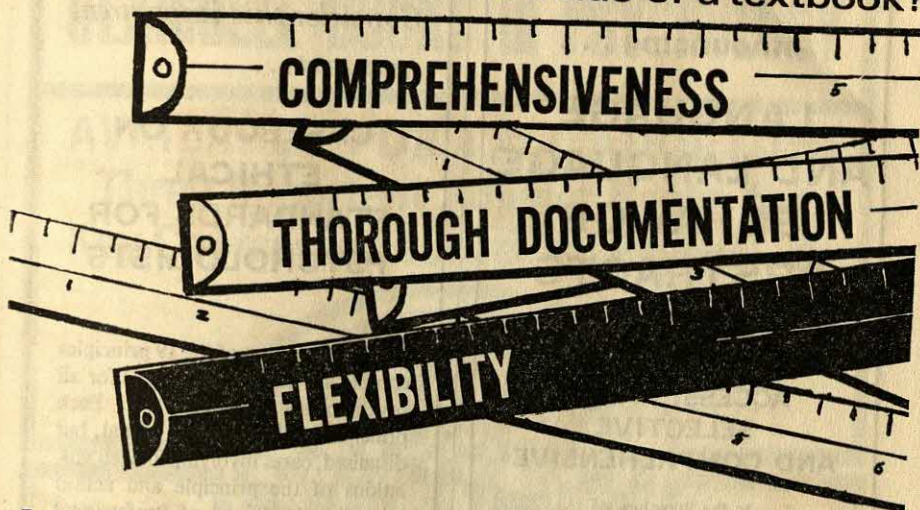
*In terms of word-attack skills. Generally, grade levels are not assigned in the program because of the uniqueness of the learning technique.

S R A

Science Research Associates, Inc.
259 E. Erie St., Chicago, Ill. 60611

A Subsidiary of IBM

How do you measure the value of a textbook?



*Does the psychology text you are presently using
measure up as well as this outstanding new textbook?*

Heckel-Peacock

TEXTBOOK OF GENERAL PSYCHOLOGY

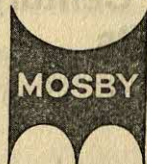
Designed for college and university courses in "General Psychology", this new text provides a broad overview of psychology at the introductory level. Comprehensive and thoroughly documented, its well-planned flexible format is sufficiently detailed to be used without supplementation for the usual one quarter or one semester course. Yet, it is flexible enough to be expanded to your specific requirements with reprints from the Bobbs-Merrill Series or from **SCIENTIFIC AMERICAN**.

Logically organized into six analytical sections, it introduces biological and developmental topics early in the presentation and emphasizes the developmental, physiological and historical areas. You will find that it provides a more expansive treatment of the *physiological, sensory, group and developmental areas* than other introductory textbooks. The easily understood discussions examine all essential aspects of modern psychology in depth with analogies to everyday behavior and experience.

A complimentary Teacher's Guide and Test Booklet is provided each instructor adopting this new book. Examine a copy of **TEXTBOOK OF GENERAL PSYCHOLOGY** and compare it to the text you are presently using. You may find that it contains *more* of the essential features you require for your introductory courses.

by **ROBERT V. HECKEL, Ph.D.**, Professor of Psychology and Director of Clinical Training, University of South Carolina, Columbia, South Carolina; and **LELON J. PEACOCK, Ph.D.**, Professor of Psychology, University of Georgia, Athens, Georgia. Contributing author: **ROBERT E. McCARTER, Ph.D.**, University of South Carolina, Columbia, South Carolina. Publication date: September, 1966. 480 pages plus FM I-XII, with 69 figures, 7" x 10". Price, \$8.50.

THE C. V. MOSBY COMPANY
3207 Washington Boulevard



Publishers
St. Louis, Mo. 63103

Announcing . . .

LANGUAGE AND LANGUAGE BEHAVIOR ABSTRACTS

*...a new journal of abstracts
whose aim is to provide*

**ACCESS: RAPID,
SELECTIVE
AND COMPREHENSIVE**

*to the literature of
language and language behavior*

- whatever the disciplinary focus
- whatever the country of origin
- whatever the language in which it is written

**... a publication of CRLLB AND
BELC**

Center for Research on Language and
Language Behavior, University of
Michigan, Ann Arbor, Michigan.

Bureau pour l'Enseignement de la Langue
et de la Civilisation Françaises a l'Etranger,
Paris, France.

Subscription price is \$22.50 a volume. Each
volume will contain four quarterly issues.
Each issue will contain 1,000 abstracts.

Subscription orders should be sent to:



Subscription Manager, LLBA

**Appleton-Century
Crofts**

Division of Meredith Publishing Co.
440 Park Ave. South, New York, N.Y. 10016

*Consider this important
new text...*

CASEBOOK ON ETHICAL STANDARDS FOR PSYCHOLOGISTS

The application of the 19 principles
comprising the Code of Ethics for all
psychologists is presented. Each
principle is illustrated by actual, but
disguised, cases involving alleged vio-
lations of the principle and action
taken by the Board of Professional
Affairs.

Four Appendices state and amplify
the additional principles applicable
to specific areas of psychology:

*—Ethical Practices in Industrial Psy-
chology*

*—Automated Test Scoring and Inter-
pretation Practices*

*—Standards for APA Directory
Listings of Private Practice*

*—Guiding Principles for the Humane
Care and Use of Animals*

The arrangement of the Casebook
into a general section of cases and
four additional specific sections in-
creases its versatility as a reference
book for individuals and as a text for
instructional purposes.

Price: \$1.00

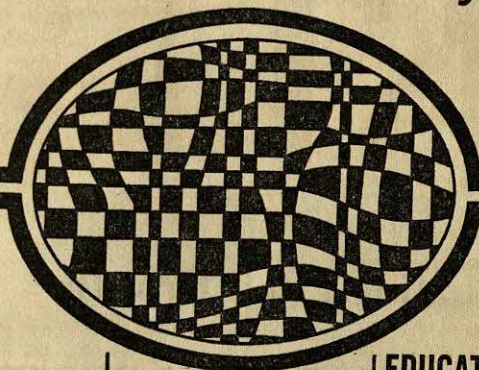
Order From:

**AMERICAN PSYCHOLOGICAL
ASSOCIATION**

1200 Seventeenth Street, N. W.
Washington, D. C. 20036

BLAISDELL BOOKS in PSYCHOLOGY

AVAILABLE FOR FALL ADOPTION
Three Important Adjunct Books for
Key Courses in Educational Psychology



READINGS IN THE PSYCHOLOGY OF CHILDHOOD AND ADOLESCENCE.

Edited by
William J. Meyer —
for courses focusing on
human development
Contains 65 theoretical
papers and research
reports dealing with
the main features of
child and adolescent
development.

STUDIES IN EDUCATIONAL PSYCHOLOGY.

Edited by
Raymond G. Kuhlen
for courses focusing on
the psychology of
school learning
Contains 50 papers
dealing with the
learning process with
special attention to the
cognitive, personality,
and motivational
factors involved.

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT: Contributions to Theory and Practice.

Edited by
David A. Payne and
Robert McMorris —
for courses focusing
on tests and
measurements
Contains 54 theoretical
and research papers
providing a broad
perspective on test
development with
an emphasis on the
assessment and
prediction of learning
outcomes.



BLAISDELL PUBLISHING COMPANY
A DIVISION OF GINN AND COMPANY
275 Wyman Street, Waltham, Massachusetts 02154

Coming Spring, 1967—

Evaluating Pupil Growth:

PRINCIPLES OF TESTS AND MEASUREMENT, THIRD EDITION

By **J. Stanley Ahmann**, Academic Vice President and Professor of Psychology, Colorado State University; and **Marvin D. Glock**, Professor of Educational Psychology, School of Education, Cornell University.

This revision of a widely acclaimed book is designed for courses in Tests and Measurements in the departments of Education, Educational Psychology, General Psychology, and Physical Education.

Thoroughly modern, the text:

Contains up-dated discussions of standardized instruments; for example, the Differential Aptitude Test.

Presents specific examples of measuring educational and behavioral objectives for considering the use of computers in text scoring, analysis, and reporting.

Includes new discussion questions and recent suggested readings.

1967. 6 x 9. Est. 500 pp. \$7.95 tent.

ALLYN AND BACON, INC.

Recently Published—

The Causes of Behavior

READINGS IN CHILD DEVELOPMENT AND EDUCATIONAL PSYCHOLOGY, SECOND EDITION

Edited by **Judy F. Rosenblith**, Wheaton College; and **Wesley Allinsmith**, University of Cincinnati.

Used by hundreds of colleges and universities in the first edition, this eminently successful book has been revised from suggestions made by its users. Approximately 40% of the articles are new. Of these new selections, approximately one-half have appeared since 1961, many as recently as 1965. The second edition features an expanded discussion of the culturally disadvantaged child, and a more thorough discussion of Piaget. A test manual is available.

1966. 8½ x 11. Paperbound. 608 pp. \$6.50

Recently Published—

An Introduction to Statistics and Measurement

A PROGRAMMED BOOK

By **Lowell A. Schoer**, State University of Iowa

The text presents statistical concepts necessary for dealing with norms, predictions, validity, reliability, regression, and the standard error of measurement.

"Mr. Schoer has done an excellent job of covering the material needed in a first course in Statistics and Measurement. His arrangement of frames is in easily understood [form], and his instructions are sufficient and clearly stated." —Dr. Earl D. Ludlow, Missouri State Teachers College

1966. 8½ x 11. Paperbound. 201 pp. \$3.95

Widely used—

Educational Psychology, Second Edition

By **James M. Sawrey** and **Charles W. Telford**, both of San Jose State College

A comprehensive presentation of the psychological foundations of education. The authors stress the principles of motivation and learning as the central concepts of educational psychology.

1964. 6 x 9. 592 pp. \$7.95

A Student Guide by **Earl L. Jandron** of San Jose State College accompanies the text. Each chapter includes class-tested programmed units. 1964. 8½ x 11. Paperbound. 184 pp. \$3.75.

Dept. Z-1, 150 Tremont St., Boston, Mass. 02111

hARPER

Just published



torchbooks

RICHARD M. JONES, Editor

Contemporary Educational Psychology *Selected Readings*

The Contents:

PART ONE: PERSPECTIVES. Introduction: Richard M. Jones: "Education in Depth" and "The New Curricula." I. Loren C. Eiseley: Man: The Lethal Factor. II. Jerome S. Bruner: Education as Social Invention. III. Aldous Huxley: Education on the Non-verbal Level. IV. Lawrence S. Kubie: The Forgotten Man of Education. V. Lawrence S. Kubie: Research in Protecting Pre-conscious Functions in Education.

PART THREE: APPLICATIONS. X. Dean M. Laux: A New Role for Teachers? XI. Donald Kingsbury: An Experiment in Higher Education. XII. Richard M. Jones: The Role of Self-Knowledge in the Educative Process. XIII. Richard M. Jones: Some Educational Aspects of Group-Leader Interaction. XIV. Elena Goldstein Werlin: An Experiment in Elementary Education. XV. Wilfrid Hamlin: Fission and Fusion. TB/1292 \$2.75

PART TWO: SYSTEMS. VI. Howard E. Gruber: Education and the Image of Man. VII. Ulrich Neisser: The Multiplicity of Thought. VIII. Jerome S. Bruner: The Course of Cognitive Growth. IX. George S. Klein: Consciousness in Psychoanalytic Theory: Some Implications for Current Research in Perception.



Complete catalog of Harper Torchbooks [557 volumes now in print] available on request from Dept. 51, HARPER & ROW, Publishers, 49 E. 33rd St., New York, N.Y. 10016

*A new revision
of major significance . . .*

APA PUBLICATION MANUAL 1967 Revision

The APA PUBLICATION MANUAL has undergone a complete revision for the first time in 10 years. Although the revised edition prescribes the same basic format for manuscripts as the previous edition, major changes and additions were made in the text in an effort to clarify and simplify the instructions on how manuscripts should be prepared.

The APA PUBLICATION MANUAL is intended primarily as a guide for authors and editors in the preparation of material to be published in APA journals. However, in preparing the original edition it was recognized that the MANUAL could, and should, be written to accommodate a wider range of manuscripts. The revised PUBLICATION MANUAL has retained that original concept.

April, 1967

\$1.50

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D. C. 20036

To be published in the Fall of 1967

Otis-Lennon Mental Ability Test

A completely new edition of the well-known
OTIS QUICK-SCORING MENTAL ABILITY TEST

In addition to retaining many of the outstanding features of previous editions, this entirely new series incorporates further refinements in item content, organization of the various levels, and scaling and norming procedures. These improvements were specifically designed to guarantee a continuous, carefully articulated scale of general mental ability.

FIVE LEVELS of the *Otis-Lennon Mental Ability Test* have been devised for efficient measurement of the range of ability commonly found in grades K-13. These are:

- Primary Level (Grades K-1.5)**
- Elementary I Level (Grades 1.5-3.9)**
- Elementary II Level (Grades 4.0-6.9)**
- Intermediate Level (Grades 7.0-9.9)**
- Advanced Level (Grades 10.0 and above)**

The two lower levels, designed for Grades K-3, contain pictorial and geometric test content sampling the mental processes of classification, following directions, quantitative reasoning, verbal reasoning, verbal conceptualization, and reasoning by analogy. No reading is required at these levels.

The upper three levels contain verbal, symbolic, quantitative, and figural test material, sampling fourteen different mental processes, measuring abstract reasoning ability.

Two general types of norms are available:

GRADE NORM GROUP — Percentile ranks and stanine scores indicate each pupil's standing in the national sample at a given grade level.

AGE NORM GROUP — The Deviation IQ Score (DIQ) indicates the pupil's relative position in the national sample of pupils of a similar chronological age. The percentile rank and stanine scores further clarify the meaning of the DIQ score.

For further information on Otis-Lennon Mental Ability Test, write:

Harcourt, Brace 
& World, Inc. TEST DEPARTMENT

757 Third Avenue, New York, N.Y. 10017

Chicago • San Francisco • Atlanta • Dallas

Now! A Self-Selection Textbook

INTRODUCTION TO GENERAL PSYCHOLOGY

Consulting Editor

Jack Vernon

Department of Otolaryngology

The following titles are part of the Wm. C. Brown INTRODUCTION TO GENERAL PSYCHOLOGY: A SELF-SELECTION TEXTBOOK. Each chapter in this textbook, standing alone, affords an introduction to one of the primary themes in the tapestry of psychology and may be used as supplementary material to a course textbook already in use. The chapters which are listed here are those which have direct applicability as supplementary material for the educational psychology course. Each chapter is published individually as a self-contained unit of approximately 50 pages in length, and punched to fit a standard three-ring binder—65 cents each.

CHAPTERS NOW AVAILABLE

PHYSIOLOGICAL BASIS OF PSYCHOLOGY by John C. Armington, Ph.D., Northeastern U.
HUMAN LEARNING by Richard Dolinsky, University of Toledo
MOTIVATION by A. K. Bartoshuk, Brown University
THE ANALYSIS OF HUMAN OPERANT BEHAVIOR by Ellen P. Reese, Mt. Holyoke College
SYMBOLIC PROCESSES by S. Glucksberg, Princeton University
SOCIAL PROCESSES by Peter Suedfeld, Rutgers
CHILD DEVELOPMENT by Charles P. Smith, Brooklyn College
PERSONALITY by Robert W. Lundin, The University of the South
PERCEPTION by R. H. Day, Monash University
EMOTIONS by Jose M. R. Delgado, Yale University
MEASUREMENT by Leonard M. Horowitz, Stanford University
THE NATURE AND SCOPE OF PSYCHOLOGY by Richard H. Henne-
man, University of Virginia

- *Chapters already being used in more than 160 schools*
- *Additional titles in preparation*
- *Examination copies available upon request*

WM. C. BROWN COMPANY PUBLISHERS

BOX 539



DUBUQUE, IOWA 52001

From Macmillan

THE PSYCHOLOGICAL FOUNDATIONS OF EDUCATION SERIES

General Editor: Victor H. Noll, Michigan State University

Intensive investigation in the field of educational psychology has resulted in the establishment of a wide variety of significant topics that warrant independent consideration. This paperbound series is designed to provide a deeper examination of specific areas which are not given thorough treatment in many textbooks now in use. The Noll series offers a convenient source of up to date supplementary material. Theoretical material is an important part of each book, but the emphasis is centered on practical applications.

PSYCHOLOGY OF ADOLESCENCE FOR TEACHERS

By Glenn Myers Blair and R. Stewart Jones, both, University of Illinois
1964, 128 pages, \$1.50

PSYCHOLOGY OF THE CHILD IN THE CLASSROOM

By Don C. Charles, Iowa State University
1964, 86 pages, \$1.25

THE PSYCHOLOGY OF LEARNING IN THE CLASSROOM

By Robert C. Craig, Michigan State University
1966, 85 pages, \$1.25

THE MENTALLY RETARDED CHILD IN THE CLASSROOM

By Marion J. Erickson, Ypsilanti Public Schools
1965, 114 pages, \$1.25

PROBLEM SOLVING IN THE CLASSROOM

By Bryce B. Hudgins, Washington University
1966, 74 pages, \$1.25

TEACHER SELF-EVALUATION

By Ray H. Simpson, University of Illinois
1966, 100 pages, \$1.25

GUIDANCE IN THE CLASSROOM

By Ruth Strang, University of Arizona, and Glyn Morris, Assistant Supervisor of Guidance and Curriculum, Lewis County, New York
1964, 118 pages, \$1.50

GIFTED CHILDREN IN THE CLASSROOM

By E. Paul Torrance, University of Minnesota
1965, 102 pages, \$1.25

Write to the faculty service desk for examination copies.

THE MACMILLAN COMPANY

866 Third Avenue, New York, N.Y. 10022

McGraw-Hill Books

Available Spring

ELEMENTARY STATISTICAL PROCEDURES

By CLINTON I. CHASE, Associate Professor of Education and Associate Director of the Bureau of Educational Studies and Testing. An introductory text in statistical methods designed primarily for students in the social sciences. Emphasis is on deriving procedures from the logical structures which underlie the methods of analyzing data.

EDUCATIONAL PSYCHOLOGY, Second Edition

By ARDEN FRANDSEN, Professor of Psychology, Utah State University. Assumes that educational psychology should give teachers a set of psychological principles which can be applied in their own teaching situations. The book is comprehensive, integrating both classic and current experiments and theories in learning. An instructor's manual is available.

PSYCHOMETRIC THEORY

By JUM C. NUNNALLY, JR., Professor of Psychology, Vanderbilt University. McGraw-Hill Series in Psychology. A comprehensive graduate-level text for courses and seminars in psychological measurement. The material is intended for the general student, yet there is a high degree of sophistication.

COUNSELING AND PSYCHOTHERAPY: An Overview

Edited by DUGALD S. ARBUCKLE, Boston University. McGraw-Hill Series in Guidance, Counseling and Student Personnel in Education. A number of authorities, practitioners and theorists bring their intelligence, experiences and beliefs to the various subjects discussed. The editor comments in a final chapter and at the end of each preceding chapter.

Examination copies upon request

McGraw-Hill



Book Company

330 West 42nd Street, New York, New York 10036

SEEKING A PUBLISHER?

Book publication is one of the foundation stones of a scholar's career in gaining recognition and advancement. Exposition Press, under its special academic imprint **Exposition University Books** offers you a complete publishing service... including editorial supervision, a dynamic marketing and promotional program, and professional distribution.

Your inquiries and manuscripts are invited (all subjects welcome). An editorial report furnished without obligation.

FREE: Two fact-filled illustrated brochures explain the behind-the-scenes story of publishing, describe our first edition plan, and present a breakdown of contract terms and typical costs.

Write Dept. 63

EXPOSITION PRESS

386 Park Ave., So., N. Y., N. Y. 10016

Emotional Disturbance and School Learning: a Book of Readings

Donald H. Clark, Hunter College
of New York City University
Gerald S. Lesser, Harvard University

"School learning is not simply learning the three R's; it is learning to master many things as diverse as arithmetic, hostility, spelling, and tenderness."

This compact anthology of 26 articles introduces the reader to current research on the problem of school learning. It stresses the value of a critical approach toward research and the scientific method and emphasizes that research provides information, not answers. The reader is encouraged to sift and weigh the various psychological approaches in the articles in order to develop his own critical faculties.

EMOTIONAL DISTURBANCE AND SCHOOL LEARNING affords excellent supplementary reading for courses in child development, educational psychology, mental hygiene, counselling and guidance, personality, and methods in education. For the school guidance counsellor it can help clarify the confusion surrounding the "scientific method" and, with its ample bibliography, serve as a source book for further investigation of emotional disturbance in the child. 294 pp., paperbound, \$3.50 list, \$2.80 net.

Write for your free examination copy of
EMOTIONAL DISTURBANCE AND
SCHOOL LEARNING.

S R A

Science Research Associates, Inc.
College Department
259 E. Erie St., Chicago, Ill. 60611

A Subsidiary of IBM

Four New Texts in Educational Psychology

EDUCATIONAL PSYCHOLOGY

The Teaching-Learning Process

*Allen Jack Edwards, University of Missouri
and Dale P. Scannell, University of Kansas*

This text concentrates on selected areas of psychology relevant to the needs of prospective teachers, with emphasis on the variables related to classroom acquisition, retention and transfer. A balance between research and theory is maintained throughout the text.

October

\$7.95

READINGS IN EDUCATIONAL PSYCHOLOGY

*Edited by Harold W. Bernard and
Wesley C. Huckins*

*Division of Continuing Education,
Oregon State System of Higher Education*

While placing great emphasis on the teacher's personality as the crucial variable in the learning process, this collection of up-to-date papers (some published in 1966) studies the basic features of educational psychology—motivation, learning, stages of development, intellectual processes, and evaluation.

May

paperbound

\$5.25

UNDERSTANDING HUMAN MOTIVATION

*Edited by Chalmers L. Stacey, Syracuse
University, and
Manfred F. DeMartino, Onondaga University*

A new revised edition of an outstanding compilation whose first edition has been a required classroom text at 69 colleges and universities. It now includes the additional concepts of homeostasis and competence, Maslow's most recent work on self-actualization and psychological health and McClelland's achievement motive.

Available

\$7.95

PERSPECTIVES IN THE EDUCATION OF DIS- ADVANTAGED CHILDREN

Milly Cowles, Editor, University of Georgia

A multi-discipline approach, this unique work presents papers by fourteen experts in anthropology, education, guidance, psychology, medicine, sociology, and social work, relating their disciplines to the education of disadvantaged children.

June

paperbound

\$4.95



THE WORLD PUBLISHING COMPANY
CLEVELAND AND NEW YORK

Selections in Psychology

New



RESEARCH IN CLINICAL ASSESSMENT

Edited by Edwin I. Megargee

Designed as an advanced text and as an aid to the professional clinician, this volume presents recent research studies from twenty American, Canadian, and British journals, covering the issues and problems of assessment in the mid-1960's. The first part of the book focuses on general problems of assessment research; the second on the validity of specific assessment techniques. The final section deals with the integration of clinical data in assessment. 702 pages; \$14.75

THEORIES OF COUNSELING AND PSYCHOTHERAPY

C. H. Patterson

Here are summaries of fifteen major theories or points of view in counseling and psychotherapy; each chapter has been reviewed by the theorist concerned. The book is organized into five sections: Rational Approaches; Learning Theory Approaches; Psychoanalytically Oriented Approaches; Perceptual-Phenomenological Approaches; and Existential Approaches. The author has supplied a concluding chapter on divergent and convergent trends. 518 pages; \$9.75

RECORDING AND ANALYZING CHILD BEHAVIOR:

With Ecological Data from an American Town

Herbert F. Wright

This text, combining parts of an earlier, out-of-print volume (*Midwest and Its Children* by R. G. Barker and H. F. Wright), represents a method for describing the situation and behavior of individuals in their natural habitats. Conceptual foundations of the method are given, and so are results of its application to children of a town in the central part of the United States. The book reintroduces field techniques, analytical procedures, and demonstrative data in a comparatively new area of investigation: ecological psychology. 290 pages; \$4.50; paper

LEARNING AND HUMAN ABILITIES:

EDUCATIONAL PSYCHOLOGY Second Edition

Herbert J. Klausmeier • William Goodwin

This comprehensive text emphasizes the concept of emerging human abilities—thus integrating the notions of growth and learning that are traditionally treated separately in other existing texts. The content and organization meet a wide range of needs of both students and professors. Course outlines have been prepared for the instructor and suggest a variety of combinations of all or parts of each of the eighteen chapters. The text is thus adaptable to courses with different emphases and time requirements, and to students with different backgrounds. 720 pages; \$8.95

Student workbook \$3.25

READINGS IN LEARNING AND HUMAN ABILITIES:

EDUCATIONAL PSYCHOLOGY

Richard E. Ripple

Intended for use with LEARNING AND HUMAN ABILITIES: EDUCATIONAL PSYCHOLOGY or with any current textbook in introductory educational psychology, this collection of forty-nine readings deals with elementary, secondary, and college levels. There is an introduction to each of the eight sections, an interpretative headnote for each reading, and a specially prepared article to help students interpret the statistics in the selections. 586 pages; \$4.95; paper

HARPER & ROW PUBLISHERS

49 East 33d St., New York 10016

**Of first importance—
the new, revised edition of a
highly significant
text . . .**

STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTS AND MANUALS

This valuable publication has been approved by the governing bodies of the American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME). The current statement constitutes a revision of two documents: (a) "Technical Recommendations for Psychological Tests and Diagnostic Techniques" which was prepared by a committee of the APA and published in March 1954, and (b) "Technical Recommendations for Achievement Tests" which was prepared by two committees of the AERA and the National Council on Measurements Used in Education (now known as the NCME) and issued by the National Education Association in 1955.

These recommended standards cover not only tests as narrowly defined, but also most published devices for diagnosis, prognosis, and evaluation. The standards apply to interest inventories, personality inventories, projective instruments and related clinical techniques, tests of aptitude or ability, and achievement tests—especially those used in an educational setting.

1966

40 pages

\$1.00

available from: AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N.W.
Washington, D.C. 20036

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Diary No. 92
Date 10.9.76
Lib
Research
Number 3

Volume 58

June 1967

Number 3

CONTENTS

<i>Critique of Cattell's "Theory of Fluid and Crystallized Intelligence: A Critical Experiment".....</i>	129
LLOYD G. HUMPHREYS	
<i>Comparability of Group Television and Individual Administration of the Peabody Picture Vocabulary Test: Implications for Screening.....</i>	137
GEORGE A. FARGO, DORIS C. CROWELL, MARY H. NOYES, ROBERT Y. FUCHIGAMI, JOHN M. GORDON, AND PETER DUNN-RANKIN	
<i>Differential Personality Development in Young Adults of Markedly Different Aptitude Levels.....</i>	141
WALTER T. PLANT AND EDWARD W. MINIUM	
<i>Learner Competence, Model Competence, and Number of Observation Trials in Vicarious Learning.....</i>	153
FREDERICK H. KANFER AND PRYSE H. DUERFELDT	
<i>Effects of Training Procedures on Creativity Test Scores.....</i>	158
DENNIS R. RIDLEY AND ROBERT C. BIRNEY	
<i>Effect of Intrinsic and Extrinsic Reinforcement Contingencies on Learner Performance.....</i>	165
HOWARD J. SULLIVAN, ROBERT L. BAKER, AND RICHARD E. SCHUTZ	
<i>Prediction of Adolescent Popularity and Rejection from Achievement and Interest Tests.....</i>	170
HERBERT HOROWITZ	
<i>Teachers' Attitudes toward Children's Behavior Revisited.....</i>	175
ALEXANDER TOLOR, WILLIAM L. SCARPETTI, AND PAUL A. LANE	
<i>Effect of Feedback from Teachers to Principals.....</i>	181
ROBERT W. DAW AND N. L. GAGE	
<i>Erratum.....</i>	180

CRITIQUE OF CATTELL'S "THEORY OF FLUID AND CRYSTALLIZED INTELLIGENCE: A CRITICAL EXPERIMENT"

LLOYD G. HUMPHREYS¹

University of Illinois

In a complexly designed factor analysis of ability and personality variables Cattell found evidence for higher-order factors representing his hypothesized fluid and crystallized abilities. This critique demonstrates that there are methodological weaknesses at every stage of the study from the selection of the variables, through the 1st-order factoring, to the 2nd-order factoring such that one can not have confidence in the results. In contrast, a simple 9×9 table of intercorrelations provides results that are in line with Cattell's theory though they are not critical in distinguishing his theory from Vernon's hierarchical model of abilities. General principles for the design of factor-analytic investigations also emerge.

Cattell (1963) has published the results of an investigation designed to test certain ideas presented earlier (Cattell, 1941) concerning the nature of intelligence. The title "Theory of Fluid and Crystallized Intelligence: A Critical Experiment" indicates confidence in the design of the study, and the conclusions drawn indicate confidence in the results. After examination of the design and of the results, however, the present writer concluded that he lacked confidence in both. The ideas which led to the research are stimulating, but they deserve a better design and a more objective methodology.

In the study in question, second-order factors, their intercorrelations, third-order factors, and their intercorrelations were presented. The factorings on which the conclusions rest are rather far removed from the basic data, the intercorrelations of the var-

iables. The techniques of factor analysis are not sufficiently objective that a reader can accept final rotated matrices without question. This is particularly true of higher-order analyses. It is essential to have test intercorrelations and the first-order factoring at hand in order to interpret adequately, let alone criticize, higher-order factoring.²

The design of the study is complex, and probably unnecessarily so. Results will be presented later in this critique which indicate that a simple, straightforward approach provides a much less ambiguous basis for interpretations. This alternative method of analysis also has the merit that it supports the relationships among the ability variables hypothesized by Cattell!

The Variables and Their Intercorrelations

The variables consisted of two forms of the Verbal, Spatial, Reasoning, and Number tests of the Thurstone Primary Mental Abilities series,

¹The writer acknowledges the support of the University Research Board of the University of Illinois in this investigation. Unfortunately, although three random normal deviates were used in the original study, no correlations or factor loadings of these variables were included in any of the tables of basic data available to the writer.

²All but one of the basic tables of data were obtained from ADI. R. B. Cattell kindly furnished the missing table.

TABLE 1
CORRELATIONAL DATA FOR QUESTIONNAIRE VARIABLES

Variable number	Form A - Form B correlation	Highest correlation with another variable			
		Form A	With	Form B	With
1	+ .36	- .31	B 4	+ .40	B 7
2	+ .28	+ .39	Vocabulary	+ .35	Number
3	+ .30	- .38	A 9	- .44	A 11
4	+ .27	+ .37	A 14	- .40	B 7
5	+ .30	- .38	A 7	- .29	B 7
6	+ .28	- .23	A 12	- .29	B 10
7	+ .43	- .38	A 5	+ .40	B 1
8	+ .31	- .43	A 11	+ .38	B 3
9	+ .37	- .38	A 3	- .28	A 5
10	+ .01	- .16	A 6	- .29	B 6
11	+ .42	- .44	B 3	- .41	A 8
12	+ .08	- .29	A 9	+ .24	B 1
13	+ .06	- .31	A 4	+ .35	B 4
14	+ .33	+ .41	B 3	- .37	B 3
Median	+ .30	.375		.36	

one form of the Word Fluency Test of the same series, one form of each subtest in the Culture Fair Intelligence Test, and two forms of each test of the Junior-Senior High School Personality Questionnaire. The personality questionnaire variables were included in accordance with Cattell's rationale of providing "hyperplane stuff." The sample consisted of 277 high-school-age boys and girls.

There are several items of interest in the intercorrelations of these variables. These are summarized in Table 1. It is seen that the median correlation between alternate forms of the personality tests is .30. Three of these correlations are not distinguishable from zero. In almost every case one form or the other of each personality variable has higher correlations with some other variable than it has with its alternate. In many cases this is true for both forms.

While it is statistically possible to have two uncorrelated variables each substantially correlated with the same factor, one hardly expects to

see this in tests that are supposed to be alternate forms of each other. Furthermore, there should be a rather clearcut explanation for the suppression effects necessary to account for such a situation. Suppression effects that are accumulated over a substantial number of variables with the correlations involved not differing appreciably from zero suggest an operation of capitalizing on chance rather than meaningful and replicable phenomena.

One variable (§23, which is A 10 of the personality questionnaire) has no correlation with anything greater than .16. Its mean correlation with 40 other variables is +.024 and the standard deviation of these correlations is .073. This is to be compared with the standard error of a zero correlation of .060. This variable is little, if anything, more than a random normal deviate. Several others are fairly close to being the same. Linn (1964, 1965) has shown that random deviates can and do introduce "noise" into the system. Far from helping to define factors, they obscure the defi-

dition of factors. Horn (1966, in press) has also shown that factors extracted from the intercorrelations of random normal deviates can be rotated to replicate rather closely factors obtained from psychological data. Placing these two findings together permits the deduction that hyperplane stuff may serve the function of allowing greater opportunity to capitalize upon chance by increasing the ratios of number of factors and number of variables to number of cases. These ratios were increased substantially in this instance. The addition of 28 variables from the personality questionnaires more than tripled the number of variables (from 13 to 41). The inclusion of these variables also was designed to provide up to 14 additional factors.

The First-Order Factoring

Twenty-two factors, based upon the intercorrelations of the 41 supposedly psychological variables plus three explicitly defined random normal deviates, were retained for rotational purposes. This number of factors was supported by two separate tests indicating 20-24 factors. Other tests, however, indicate many fewer factors. (a) There is a marked drop in the size of the centroid loadings between Factors 10 and 11, on the one hand, and Factors 12 and 13, on the other. (b) The product of the largest loadings on the latter two factors is only about .04. In comparison the standard error of a zero correlation coefficient is .06. (c) Variable 23 (A 10), which is hardly more than a random normal deviate, has factor loadings on all factors that are appreciably larger than those of other variables with Factor 12 or any factor thereafter. (d) The first instance in which parallel forms of the Primary Mental Abilities tests depart

from each other in sign of factor loadings is on Factor 11. On Factor 17 and thereafter there is always one case in which the parallel forms of the ability variables differ from each other in sign. On the basis of these four lines of evidence, 22 factors appears to be excessive.

The First-Order Rotations

The evidence for overextraction of factors obtained from analysis of the unrotated loadings is supported by analysis of the rotated factors. There are three personality factors (XVIII, XX, and XXI), for example, defined by the three pairs of alternate forms that correlate with each other .01, .06, and .08, respectively. There are also separate factors for the single forms of each subtest of the Culture Fair test and for the Fluency test from the PMA. The reference vector correlations of other variables that supposedly convert specific factors into nonspecifics simply do not make sense. One form of a personality test may have a reference vector correlation greater than .20 with one of these specifics, but not its alternate form. Some of these correlations are relatively large, for example, the reference vector correlation of .45 of Variable 33 (B 6) on the so-called Fluency Factor on which Word Fluency has a correlation of .39, but it still seems reasonable to conclude that a specific factor was converted into a nonspecific by capitalizing on chance. First-order factors were needed for these variables for the second-order analysis even though no provision had been made for them in the design of the study and rotations were accomplished to achieve this.

One last line of evidence indicates overfactoring and contrived rotations. In showing that factors from random data could be rotated into

positions that were highly similar with those from psychological data, Horn (in press) computed squared multiple correlations between each rotated factor and all of the remaining factors. These were quite high with most being greater than .90. This may be the almost inevitable result of the process of capitalizing on chance to obtain a theoretical structure in the data. In the present study the median multiple correlations squared for the ability factors is .56 with a range from .32 to .71. For the personality factors the median is .60 with a range from .35 to .83. These latter particularly are higher than anticipated from previous work with personality tests and approach the values Horn obtained for the random data factors.

If we subtract five specific factors and three personality factors defined by variables essentially uncorrelated with each other, we are left with 14 possibly meaningful factors. This estimate does not, of course, allow for the possibility that there may not be 10 factors remaining in the personality questionnaire data and is thus a maximum figure.

The Hyperplane Count Criterion

The preceding analyses of the first-order rotations indicates that hyperplane count is an inadequate criterion of the validity of rotations. This inadequacy is brought out more clearly by considering two extreme cases in which it might be applied. A set of intercorrelations showing Spearman's hierarchical order will define only one factor with no variable having a loading of zero. This factor is uniquely determined though there is no hyperplane. In a more complex table of intercorrelations, on the other hand, if an investigator were to extract as many factors as variables,

he would find $n - 1$ variables in each of the hyperplanes. Furthermore, the percentage of variables in the hyperplane would increase directly with n in this second case. Whenever intercorrelations are overfactored, this second case is approached and the hyperplane count will be high. The inclusion of five specifics and three factors determined by zero correlations between parallel forms insures a high hyperplane count overall in the Cattell first-order analysis.

Hyperplane count is also increased by starting the analysis with many variables of low communality. Yet Tucker (1964) has shown that size of communality is one of the most critical variables in determining whether a known factor structure can be recovered from correlational matrices. Under carefully controlled conditions, including adequate communalities of the variables and adequate criteria for number of factors, hyperplane count is useful. Otherwise it provides a rationale for overfactoring, use of variables of low communality, and subsequent capitalization on chance in rotations.

The Second-Order Factoring

Eight factors were extracted in the second order. Since too many factors were rotated in the first order, and since the rotations built up a high degree of dependency among the first-order factors as a whole, it is not surprising to find that the second-order factors tend to make a larger contribution to variance than many of the first-order factors. These factors accordingly determine high row sums of squares (estimated communalities) which are in line with the high squared multiple correlations among the first-order factors. The noniterated communalities have a

median of .61 and a range from .32 to .89. Four are in the high eighties.

Of the eight factors extracted, probably only the eighth is too small to retain for rotations and interpretation. Note that this does not mean there are at least seven second-order factors in any psychological definition of the term. Rather, the high level dependencies in the first order *created* second-order factors. By the same token, the second-order factors in these data make several of the first-order factors superfluous; that is, for all practical purposes the second-order factors completely describe several of the first-order factors.

The Second-Order Rotations

There are difficulties in the second-order rotations that are ignored in the discussion of the results. After rotation there are two factors that are not interpreted. Both have substantial loadings for a number of variables. Both make a bigger contribution to variance after rotation than do most of the unrotated factors. Such factors could have provided freedom to rotate the remaining six to agree with hypotheses.

Interpretation of factors and factor loadings is of the "broad brush" variety. Loadings not in line with hypotheses are ignored. For example, ego strength, with a loading of .21, is listed as one of the defining variables for the fluid-ability factor, but *parmia*, also with a loading of .21, is not mentioned. In interpreting the crystallized-ability factor, on the other hand, no mention is made of cyclothymia or of excitability with loadings of .52 and $-.44$, respectively. Similarly, the loading of $-.52$ of the supposed first-order fluency factor on the second-order extroversion factor is not discussed. The assertion that there is an interaction

between the control factor and crystallized ability apparently depends on a correlation between the factors of .28 whereas the corresponding correlation with fluid ability is .21. Even if factor interpretation were viewed as hypothesis formation only, rather than hypothesis testing, such selectivity could not be defended.

An Alternative Analysis

Cattell structured his analysis of the data as a series of factorings in several orders in order to avoid "distortions from the unreliabilities and imperfect validities of actual scales [1953, p. 13]." This is theoretically sound, but depends in practice upon an objective factor-analytic methodology. It is useful to look at a simpler approach and to gauge the extent to which unreliability and imperfect validity mask relationships versus the extent to which a simpler design reduces computations and interpretational difficulties.

The present writer computed intercorrelations between sums of the parallel forms of the ability variables and the single-form ability tests. The result was a nine-variable correlational matrix which is presented as Table 2. One does not need to factor this matrix in order to test the hypothesis that the Cattell tests (Variables 6-9) and the Spatial test will separate from the remaining Primary Mental Ability tests. Appropriate median correlations are as follows: among the PMA without Spatial Relations, .44; among the Cattell tests and Spatial Relations, .305; between the two clusters, .255.

A factoring is, of course, readily accomplished. Communalities were first iterated for each of two, three, and four factors. Three sets of factor loadings were then computed for two, three, and four factors using the

TABLE 2
INTERCORRELATIONS OF NINE ABILITY VARIABLES

Variable	1	2	3	4	5	6	7	8	9
1. PMA-Verbal		33	46	37	46	31	23	34	25
2. PMA-Spatial	33		26	20	21	29	32	28	29
3. PMA-Reasoning	46	26		45	37	28	27	38	24
4. PMA-Numerical	37	20	45		43	20	17	28	17
5. PMA-Fluency	46	21	37	43		31	16	30	13
6. Series	31	29	28	20	31		33	45	29
7. Classification	23	32	27	17	16	33		44	28
8. Matrices	34	28	38	28	30	45	44		33
9. Topology	25	29	24	17	13	29	28	33	

Note.—All decimals omitted.

appropriate stabilized communalities. The roots for these three analyses are shown in Table 3. Loadings on two, three, and four factors are shown in Table 4.

There is always room for argument among factor analysts about the proper number of factors to rotate. The present critic is very willing to accept two factors, although there is obviously a large general factor in these data. It would also seem necessary, if one goes beyond two, to require four since there is so little difference between the third and fourth. The latter factors, in any event, do not constitute very important contributions to variance even if replicable from the sampling point of view. A plot of two factors, which is so easy to visualize as not to require a figure, confirms the separation of the Primary Mental Ability tests other than the Spatial Test from the Culture Fair tests. Furthermore, if one uses the same criteria for rotation as Cattell used in his second-order analysis; that is, allowing negligible negative loadings of the ability tests on ability factors, one obtains a correlation between the present first-order factors of .44. By allowing somewhat larger negative values, a better fit is obtained in the present nine-variable plot. This alternative

rotation produces a correlation between the factors of .57. These correlations are to be compared with a correlation reported by Cattell of .47. One concludes that there is little evidence for the effects of unreliability and imperfect validities that Cattell feared. In addition, the hypothesized result is reached with less expense, less labor, and with much greater confidence.

The congruence of observations with hypothesis does not rule out other theoretical possibilities. Vernon (1950) would have predicted the same result on the basis of his interpretations of hierarchical factors. On the basis of these data, no matter how analyzed, the two factors might

TABLE 3
ROOTS OBTAINED AFTER ITERATING COMMUNALITIES FOR TWO, THREE, AND FOUR FACTORS

Factor number	Number of factors in iterations		
	2	3	4
I	2.83	2.87	2.90
II	.59	.62	.65
III	.13	.22	.22
IV	.12	.13	.19
V	.02	.04	.06
VI	-.00	.02	.03
VII		-.03	.00
VIII			-.04

TABLE 4
TWO-, THREE-, AND FOUR-FACTOR SOLUTIONS FOR THE ABILITY VARIABLES

Variable	Two			Three				Four				
	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
1	64	-20	45	65	-22	-14	49	64	-20	-13	10	48
2	48	15	25	50	17	-32	38	49	17	-29	14	37
3	63	-19	43	63	-19	02	43	64	-19	-08	-24	51
4	54	-34	40	54	-34	05	41	54	-33	00	-15	42
5	57	-33	43	57	-34	06	44	59	-38	15	21	55
6	56	20	35	55	19	06	35	56	21	15	16	41
7	51	35	38	50	33	01	36	50	34	-01	-06	37
8	65	24	48	67	27	27	60	66	27	24	-11	58
9	44	24	25	44	23	-11	26	44	24	-12	-01	26

Note.—All decimals omitted.

as well be called intellectual-educational, rather than crystallized, and spatial-practical, rather than fluid. The correlation between the two factors determines a higher-order factor also, which Vernon would call general intelligence. Cattell's discussion of the higher-order factor, on the other hand, is ambiguous. He calls both fluid and crystallized abilities general abilities, when they are obviously not general by any accepted definition of the term. The only general-ability factor by the usual definition is the second-order factor in the writer's analysis and the third-order factor in Cattell's analysis. The general factor is determined in both analyses by the correlation between the fluid- and crystallized-ability factors.

More recent work by Horn and Cattell (1966), appearing since the above was written, is based upon a simpler design; the factors with which the authors are concerned are in the first rather than in the second order. Since factor score estimates were intercorrelated, however, these first-order factors are the equivalent of the usual second-order factors. There is still one debatable methodological issue; the present writer would have rotated and interpreted

a smaller number of factors. As a consequence of their choice of number of factors, they describe four general-ability factors, or in the more usual terms four broad group or second-order ability factors, rather than two. These are G_v (visualization) and G_s (speed) in addition to the earlier fluid and crystallized abilities.

The separation in the second order of the several first-order visualization factors from the fluid-ability factors is unexpected from Vernon's (1950) position, but the results must be considered highly tentative. The authors depended upon a commonly misused criterion for the number of factors that Horn had previously (1965) and cogently criticized and for which he had recommended a correction. Even cursory inspection reveals the presence of surplus factors. The variables with the highest estimated communalities are among those with the lowest levels of correlations with the remaining variables. Two rotated factors are specifics. A third factor is defined by four variables whose intercorrelations have a mean absolute value of .135.

Reanalysis of the Horn-Cattell data, omitting "noisy" variables 24 and 31 whose correlations with other

variables are little more than random, and iterating communalities for four, five, and six factors, indicates that four factors should be retained for rotational purposes. One of these is an abstract and visual reasoning factor which combines the Horn-Cattell fluid and visualization abilities. A second is generally comparable to their crystallized-ability factor. A third is most nearly comparable to their general speed or test motivation factor, but with the fluency variables added; intellectual speed might be a suitable name. The fourth factor in this solution is a personality factor similar to the Horn-Cattell PSI.

If a fifth factor is retained for rotations, fluid and visualization factors very similar to the ones described by Horn and Cattell emerge. There is no suggestion, as in the earlier study, that rotations were determined by the author's expectations. If five factors can be supported, then the separation of second-order visualization and fluid abilities can be supported. In addition to the fact that other criteria indicated the tenuous nature of the fifth factor, however, the estimated communality of Variable 29 climbs rapidly in 50 iterations to the unreasonably high figure of .93 and is still changing by .01. Communality estimates that are psychologically highly improbable, even though they do not exceed unity, suggest that an error has been made in selecting the number of factors. In contrast, estimated communalities stabilize quite quickly and reasonably in a four-factor solution.

One matter is somewhat clarified in the recent article (Horn & Cattell, 1966). The mechanical information factor appears to be more closely re-

lated to verbal-intellectual factors than to the spatial-practical as Vernon (1950) would expect. This is seen both in the published intercorrelations and in the writer's four-factor solution. Both of these bases for inference, however, show that the differences are not large. The writer's hunch is that neither the Vernon hierarchy nor the Cattell theory will accurately describe the relationships of the mechanical information factor to other factors when more dependable data become available.

REFERENCES

- CATTELL, R. B. Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 1941, **38**, 592.
- CATTELL, R. B. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 1963, **54**, 1-22.
- HORN, J. L. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 1965, **30**, 179-185.
- HORN, J. L. A note on the use of random variables in factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1966, **19**, 127-129.
- HORN, J. L. On subjectivity in factor analysis. *Educational and Psychological Measurement*, in press.
- HORN, J. L., & CATTELL, R. B. Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology*, 1966, **57**, 253-270.
- LINN, R. L. Use of random normal deviates to determine the number of factors to extract in factor analysis. Unpublished master's thesis, University of Illinois, 1964.
- LINN, R. L. A Monte Carlo approach to the number of factors problem. Unpublished doctoral dissertation, University of Illinois, 1965.
- TUCKER, L. R. Recovery of factors from simulated data. Paper read at *Psychonomic Society*, Niagara Falls, Ontario, 1964.
- VERNON, P. E. *The structure of human abilities*. New York, Wiley, 1950.

(Received August 5, 1966)

COMPARABILITY OF GROUP TELEVISION AND INDIVIDUAL ADMINISTRATION OF THE PEABODY PICTURE VOCABULARY TEST:

IMPLICATIONS FOR SCREENING

GEORGE A. FARGO, DORIS C. CROWELL, MARY H. NOYES, ROBERT Y.
FUCHIGAMI, JOHN M. GORDON, AND PETER DUNN-RANKIN

University of Hawaii

In a pilot study designed to test the efficiency of administering the Peabody Picture Vocabulary Test (PPVT) by means of educational television, 126 3rd- to 5th-grade children at the University of Hawaii Elementary School participated in a counterbalanced design study. All children were administered the PPVT individually and in a TV group session with $\frac{1}{2}$ Ss having the individual test first and $\frac{1}{2}$ the group test first; Forms A and B were similarly alternated for order and type of administration. An analysis of variance of the raw scores indicates the test results for individual and group TV administration are comparable. Further studies using a broader sampling group would demonstrate the feasibility of the use of educational TV as a screening medium.

This study was conducted to examine the feasibility of adapting the Peabody Picture Vocabulary Test (PPVT, Dunn, 1959) for group administration by means of educational television. The objective was to test the hypothesis that scores obtained in group TV administration would not differ significantly from those obtained in individual administration. The investigators believed that if the two administrations were found to be comparable, the economical group administration could be used to screen children and identify those who need further individual study.

The Peabody Picture Vocabulary Test was designed to measure verbal intelligence through measuring receptive vocabulary. The test provides standard score equivalents from age 2 years 3 months to 18 years 5 months for mental age, intelligent quotient, and percentile. The two forms A and B were standardized on over 4000 subjects (Ss) and have been adequately demonstrated to be equivalent (Dunn, 1965).

RELATED STUDIES

Norris, Hottel, and Brooks (1960) explored the group administration of the PPVT utilizing photographic slides projected onto a screen with an examiner providing the verbal cues. Their study demonstrated that the average scores of fifth-grade children of normal intelligence and in proper school grade for their age were not a function of the form or method of administration, that is, individual or group. Tempero and Ivanoff (1960) administered both forms of the PPVT to 150 seventh-grade students using an opaque projector and obtained a correlation of .75, which demonstrated the equivalence of IQ scores on Forms A and B. In addition, Dunn (1965, Table 1) cited nine other studies which support the equivalence of the two forms.

Although Norris and Tempero administered the PPVT to groups, they did not utilize TV administration. A similar procedure was employed by Curtis, King, and Kropp (1963) as part of a television testing study. Sup-

ported by the statements of Cronbach (1960) and Cardell (1962) indicating the value of and need for group screening, the investigators explored the possibility of television as a screening device.

METHOD

Experimental Design

A counterbalanced Treatment $\times$ Ss design (Lindquist, 1953) was utilized. Half the Ss had the individual presentation prior to the group video tape presentation and half had the group session first. There was no significant difference in the mean IQ's of the groups as measured by other individual intelligence tests. Parallel PPVT Forms A and B were used in both administrations. Table 1 presents the plan of administration.

The Treatment $\times$ S design increased the precision of the experiment by eliminating inter-S differences as an error factor. In addition, the design enabled the investigators to identify the effects of learning on the re-test. Since the form and order of administration (group or individual) were rotated for all Ss any systematic increase in test scores obtained during the second presentation could be attributed to learning.

Subjects

One hundred thirty-five children were selected from the University of Hawaii Elementary School. All third, fourth, and fifth grades participated in the study. Subsequently, the number of Ss was reduced to

126; nine Ss were eliminated as they were unavailable for the total testing sequence.

Individually administered standardized intelligence tests placed these children within an IQ range of 91-152 (with a mean of 123). It is apparent that this was a highly selected group of children. Although all ethnic backgrounds found in the state of Hawaii were represented, every child was English speaking.

Procedure

Two video tapes, one each for Forms A and B were made with modifications for use with IBM answer sheets (IBM form I.T.S. 1000A 309). An experienced psychological examiner administered the PPVT on both tapes. Two testing rooms were set up for simultaneous viewing of the TV tapes. Each room contained two television monitors (18 viewing positions to each monitor).

Prior to the TV presentation the teacher announced briefly that the students were going to take a test on TV and she asked only that the students write their names on the answer sheets. The original directions from the manual were adapted for group TV administration. The adaptation included an orientation to the task and an orientation to the use of the answer sheet. The total elapsed tape time for introduction, instructional procedures, and the test items was 30 minutes.

Plates 40 through 120 were included. These limits were defined by the tenth percentile for the third graders (or equivalent age) to the 90th percentile for the fifth graders (or equivalent age) as estimated from the norms cited in the manual. After the plate number, each word was presented twice. Items 40 through 59 were paced at 10-second intervals, 60-89 at 12-second intervals, and 90-120 at 15-second intervals. At specified intervals throughout the tape, the examiner reminded the children to look at all four pictures and pointed out the increasing difficulty of the items, encouraging them to respond as best they could. The Ss marked their responses on the IBM answer sheets.

Individual presentations were administered to each S as previously described following the PPVT standardized procedure (Dunn, 1965).

RESULTS AND CONCLUSION

An analysis of variance of the scores obtained on the 126 individual and group TV presentations yielded a Be-

TABLE 1
DESIGN FOR TEST ADMINISTRATION

Administration	Individual	TV
First		
Form A	Group 1 <i>N</i> = 29	Group 3 <i>N</i> = 30
Form B	Group 2 <i>N</i> = 34	Group 4 <i>N</i> = 33
Second		
Form A	Group 4 <i>N</i> = 33	Group 2 <i>N</i> = 34
Form B	Group 3 <i>N</i> = 30	Group 1 <i>N</i> = 29

tween Administration F ratio of .75. Since the F ratio is less than the critical value, the variation in the data may be attributed to chance and the null hypothesis was not rejected.

Three t tests were applied to the scores comparing performance on Individual Administration versus Group Administration, scores on Form A versus Form B, and scores on First Administration versus Second Administration. In no case did the differences in means reach significance (Table 2). These findings confirm those of Norris et al. that on retest a one-point gain may be attributed to a practice effect. The F test showed no significant difference at the .01 level in variance when scores were compared by Form or Order of Administration but differences slightly exceeded significance when Individual and Group Administration scores were compared.

DISCUSSION

Although Ss used represent a rather select group in terms of intellectual ability, they function over a wide enough range to support the appropriateness of TV as a medium for group screening. The apparent comparability in scores obtained under the two types of test administration points to the feasibility of the use of the TV administration of the PPVT for group testing. The greater variability in the TV group administration can probably be minimized by more careful monitoring of the testing situation and improving the viewing conditions with spaced desk placement and optimal lighting for the TV screen.

There are several important implications in group TV test administration:

1. TV screening is more economical in time and personnel. The cost in time

TABLE 2
PPVT PERFORMANCE OF 126 THIRD-,
FOURTH-, AND FIFTH-GRADE STUDENTS
(RAW SCORES)

	Form		Administration		Order	
	A	B	Individual	TV Group	1st	2nd
M	82.91	82.67	83.11	82.47	82.53	83.06
SD	8.97	10.89	8.77	11.04	9.91	9.95
t^a	0.23		.060		.044	
F^b	1.47*		1.59**		1.009	

^a $N = 120$.

^b $df = 125, 125$.

* $p < .05$.

** $p < .01$.

of testing 100 children on the PPVT individually would be 25 examiner hours. Using TV administration the same task can be accomplished with $2\frac{1}{2}$ man-hours.

2. The class is not repeatedly interrupted by the removal of each child and the test may be taken in the classroom.

3. The most proficient and effective test administrator with the best enunciation can be selected.

4. A taped presentation provides consistency and high reliability between repeated administrations.

5. The screening test can be presented simultaneously throughout the school district to obtain longitudinal results that may be used for inter-school comparisons.

6. The use of IBM answer sheets adds speed and accuracy in scoring, in addition to reducing the per-pupil cost of testing and scoring.

Educational television has been used successfully for teaching and demonstration purposes. The present pilot study demonstrates its use can be extended to evaluation as a group-

screening medium. An appropriate next step would be a replication of this study with a broader population sample.

REFERENCES

- CARDALL, A. J. *Cardall test of practical judgment*. Yardley, Pa.: Cardall Associates, 1962.
- CRONBACH, L. J. *Essentials of psychological testing*. (2nd ed.) New York: Harper, 1960.
- CURTIS, H. A., KING, F. J., & KROPP, R. P. Validity studies of scores from a delinquency proneness scale. *Psychological Reports*, 1963, **12**, 271-278.
- DUNN, L. M. *Peabody Picture Vocabulary Test*. Nashville: American Guidance Service, 1959. (Manual)
- DUNN, L. M. *Peabody Picture Vocabulary Test*. Minneapolis: American Guidance Service, 1965. (Manual)
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.
- NORRIS, R. C., HOTTEL, J. V., & BROOKS, S. Comparability of Peabody Picture Vocabulary Test scores under group and individual administration. *Journal of Educational Psychology*, 1960, **51**, 87-91.
- TEMPERO, H., & IVANOFF, J. *Effectiveness of the P.P.V.T. with seventh grade pupils*. Lincoln: University of Nebraska Press, 1960.

(Received May 23, 1966)

DIFFERENTIAL PERSONALITY DEVELOPMENT IN YOUNG ADULTS OF MARKEDLY DIFFERENT APTITUDE LEVELS¹

WALTER T. PLANT AND EDWARD W. MINIUM
San Jose State College

Differential personality changes for high- and low-aptitude groups were studied. Personality test and retest data were utilized from 5 different longitudinal studies; 3 studies were over a 2-yr. time period and 2 were over 4 yrs. Personality changes over 2 and 4 yrs. were studied for young adult males and females separately. 40 analyses of covariance were reported with retest means adjusted for initial test score. There was substantial evidence for: (a) young adults of higher aptitude to exhibit more personality change over time and in the direction of the trend of college students in general, and (b) young adults of higher aptitude to exhibit more "psychologically positive" personality development over time.

Contrary to the widely read and quoted conclusions of Jacob (1957), all reported studies of personality change in college students during college indicate that measurable changes occur. From the earliest studies (Corey, 1936, 1940; Kuhlén, 1941; Newcomb, 1943) to the last published study (Lehmann, Sinha, & Hartnett, 1966), and reviews on the topic (Bloom & Webster, 1960; McCullers & Plant, 1964; Webster, Freedman, & Heist, 1962), changes in attitude, interest, personality, and value test scores have been reported for college students during their college years. With respect to certain of the variables with which the present study is concerned, it has been consistently reported from institutions as different as Michigan State University (Lehmann, 1963; Lehmann & Dressel, 1962, 1963; Lehmann et al., 1966), San Jose State College (Plant, 1958a, 1958b, 1962, 1965), University of

Santa Clara (Foster, Stanek, & Krasowski, 1961), Vassar College (Freedman, 1960; Webster, 1958; Webster et al., 1962), and six California junior colleges (Telford & Plant, 1963) that:

1. Freshman students have significantly higher Authoritarianism, Dogmatism, or Ethnocentrism scale means than do upperclass or senior students; and

2. Freshman students have significantly higher Authoritarianism, Dogmatism, or Ethnocentrism scale means than they do when retested at the end of their sophomore, junior, or senior years.

The interpretation of these reported results has generally been that the college experience has been a "liberalizing" one in terms of personality development. Either implicitly or explicitly, the reported personality changes have been attributed to the college impact upon student personality.

There have been only four studies published to date on personality changes associated with college attendance in which comparison groups have been included. In these studies, groups of subjects (Ss) were studied over a 2-year period (Plant, 1958b, 1962, 1965; Telford & Plant, 1963)

¹ Research conducted under Contract S-042, Cooperative Research Branch, United States Office of Education, Department of Health, Education and Welfare. Appreciation is herein expressed to Robert B. Clarke and Allan Pratt for statistical consultation and computer analyses respectively.

and over a 4-year period (Lehmann & Dressel, 1963; Plant, 1962, 1965). Comparison was made between Ss in college attendance throughout the study period and Ss with no college experience beyond matriculation or some amount less than the study period. In the several studies changes were found to be similar for both types of Ss. These 2- and 4-year results have called the whole literature dealing with the effects of a collegiate experience upon student personality into question.

It may very well be that what investigators have been reporting as changes in personality characteristics resulting from college attendance are developmental changes in personality characteristics for bright young adults irrespective of their higher educational attainment during a given period of time. Longitudinal investigations of the subsequent intellectual development of initially brighter-than-average Ss indicate that changes continue well into adulthood (Bayley, 1955, 1957; Bayley & Oden, 1955; Owens, 1953) and to an extent not characteristic of less talented Ss. The current study attempted to determine if non-intellectual characteristics change over time for brighter-than-average Ss in the same fashion.

The pool from which the study groups were obtained was presumed to be comprised of brighter-than-average Ss. All Ss applied for admission to a college, were admitted, and the majority attended college. Evidence relevant to support of the presumption has been reviewed by Cronbach (1960), and reported in local studies (Conry & Plant, 1965; Plant & Lynd, 1959; Plant & Richardson, 1958).

METHOD

All data needed for the current studies were available from three separate longitudinal research projects dealing with "college impact," each of which incorporated comparison groups as a part of the research

design. There were five separate studies undertaken: three with a 2-year time interval between testing and retesting with the non-intellectual measures, and two with a 4-year interval. A description of the samples, instruments used, and statistical analyses performed is found below for each of the five studies. Because statistically significant sex differences were found for the study variables, all analyses were made separately for males and females. For purposes of the present study, no differentiation was made between Ss who were enrolled for the entire period under study and those who were in attendance for a lesser time.

Original Samples and Instruments Used

Brief descriptions of the total samples from which the current study samples were obtained are presented below. Abbreviations for the several scales which were used follow the full names of the tests, and will be used subsequently in reporting the results.

Study 1. Two-year interval between testing and retesting. Original source of data: Plant (1958b).

A group of 755 Ss (356 males and 399 females) was tested in 1953 with a 32-item California Ethnocentrism Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), or E Scale, and the 1949 form of the American Council on Education Psychological Examination for college freshmen or ACE as a part of the college prematriculation testing program at San Jose State College. The Ss were retested in 1955 with the E Scale.

Study 2. Two-year interval between testing and retesting. Original source of data: Plant (1962, 1965).

A group of 1,440 Ss (600 males and 840 females) was tested in 1958 with a 30-item California Ethnocentrism Scale (Adorno et al., 1950) or E Scale, the Dogmatism Scale (Rokeach, 1956) or D Scale, the Gough revision of the Authoritarianism Scale (Gough, 1951) or F Scale, and the 1949 form of the ACE as a part of the college prematriculation testing program at San Jose State College. The Ss were retested in 1960 with the E, D, and F Scales.

Study 3. Two-year interval between testing and retesting. Original source of data: Telford and Plant (1963).

A group of 1,678 Ss (926 males and 752 females) was tested in 1960 with the Dogmatism Scale (Rokeach, 1956) or D Scale, Allport-Vernon-Lindzey Study of Values (Allport, Vernon, & Lindzey, 1960) or AVL, the Sociability: Sy, Self-Control: Sc,

Achievement via Independence: Ai, Intellectual Efficiency: Ie, and Responsibility: Re scales from the California Psychological Inventory or CPI (Gough, 1957), and Form A of the School and College Ability Test or SCAT as a part of the prematriculation testing program in six public junior colleges in California: Coalinga, Contra Costa, Foothill, Hartnell, Monterey Peninsula, and San Jose City colleges. The *Ss* were retested with the D Scale, the AVL, and the five scales from the CPI in 1962.

Study 4. Four-year interval between testing and retesting. Original source of data: Plant (1958a).

A group of 271 *Ss* (137 males and 134 females) was tested in 1953 with a 32-item Ethnocentrism Scale (Adorno et al., 1950) or E Scale, and the 1949 form of the ACE as a part of the college prematriculation testing program at San Jose State College. The *Ss* were retested in 1957 with the E Scale after they had finished college. This was a special study in which no attempt was made to retest original *Ss* who failed, for any reason, to finish the B.A. degree in the 4-year period.

Study 5. Four-year interval between testing and retesting. Original source of data: Plant (1962, 1965).

A group of 1,032 *Ss* (448 males and 584 females) was tested in 1958 with a 30-item California Ethnocentrism Scale (Adorno et al., 1950) or E Scale, the Dogmatism Scale (Rokeach, 1956) or D Scale, the Gough revision of the Authoritarianism Scale (Gough, 1951) or F Scale, and the 1949 form of the ACE as a part of the college prematriculation testing program at San Jose State College. The *Ss* were retested in 1962 with the E, D, and F Scales.

It was felt that five separate studies which differed somewhat in the variables utilized, in the types of *Ss* studied, and in the time periods studied would enhance the generality of findings regarding change in personality as related to ability.

Statistical Analyses Performed

Within each subgroup of males and of females in each of the five studies, those *Ss* were selected whose ACE or SCAT total raw scores were in the upper quarter or in the lower quarter of the distribution for the intact subgroup. The resulting subgroups will henceforth be referred to as "high-aptitude" and "low-aptitude" groups. These groups, so constituted, became the object of the present analyses.

To examine the possibility of differential changes in nonintellectual scores over the

TABLE 1
SAMPLE SIZE AND APTITUDE SCORES FOR
ORIGINAL GROUPS FROM WHICH
SUBJECTS WERE SELECTED FOR
STUDY

Subjects	Study number				
	1 ACE total	2 ACE total	3 SCAT total	4 ACE total	5 ACE total
Males					
<i>N</i>	356	600	926	137	448
<i>M</i>	103.2	108.5	63.7	106.8	109.3
<i>SD</i>	23.0	21.1	17.0	21.6	21.5
Females					
<i>N</i>	399	840	752	134	584
<i>M</i>	100.2	106.4	58.2	103.3	107.0
<i>SD</i>	23.3	21.1	17.4	18.4	21.1

2- and 4-year periods as a function of level of aptitude, analyses of covariance were performed. Initial scores on each nonintellectual variable were used to adjust final scores; the tests of significance were therefore made on the adjusted retest means between the high- and low-aptitude groups. Summed across the five studies one finds 20 retest variables, and since the analyses were performed separately for male and female groups, 40 such tests resulted.

High- versus Low-Aptitude Groups Studied

The number of cases in the original samples from which *Ss* were selected, and data relative to academic aptitude scores of these groups, are presented in Table 1. In the high- and low-aptitude groups as constituted for the present study, the number of cases in each approximated one-quarter of the size of each group reported in Table 1. Occasionally the number is less since in the subsequent analyses it was discovered that for some variables in some studies, no retest score was available on one or more of the nonintellectual measures. In approximately half of the groups, no scores were missing. In the remainder, from one to 15 scores were missing, and these *Ss* were necessarily eliminated from the present study.

RESULTS

Preliminary Considerations

The data describing performance on the nonintellectual variables for the high- and low-aptitude subgroups are

TABLE 2

COMPARISONS OF PERSONALITY TEST SCORE CHANGES OVER TWO YEARS FOR HIGH- AND LOW-APTITUDE MALES: STUDIES 1, 2, AND 3

Test variable	Aptitude group	Mx^a	My^b	SDy	$My\ adj^c$	$\Delta\ cov^d$	F_M^e	N
Study 1 E Scale	High	85.5	77.4	22.2	79.9	—	1.61	89
	Low	93.9	86.2	26.0	83.6			89
Study 2 D Scale	High	153.0	143.3	24.7	147.5	—	.14	150
	Low	166.8	152.7	26.1	148.5			150
E Scale	High	80.6	73.5	22.0	78.1	—	3.45	150
	Low	93.9	86.7	23.8	82.1			149
F Scale	High	117.3	101.5	21.7	105.9	—	9.73**	150
	Low	131.9	117.3	21.6	112.9			149
Study 3 CPI: Sy	High	24.4	25.0	5.2	24.5	+	.11	231
	Low	22.7	23.9	5.0	24.4			229
CPI: Sc	High	25.6	26.8	7.6	26.8	—	.49	231
	Low	25.7	27.2	8.2	27.2			229
CPI: Ai	High	19.6	21.6	4.2	20.1	+	20.09**	231
	Low	14.7	16.8	4.8	18.3			229
CPI: Ie	High	38.1	39.5	5.3	37.9	+	9.90**	231
	Low	31.3	34.7	5.8	36.3			229
CPI: Re	High	30.3	31.3	4.7	30.1	+	5.05*	231
	Low	26.6	28.2	4.8	29.3			229
AVL: Theor	High	46.1	46.5	7.4	46.0	+	26.33**	231
	Low	43.8	42.5	6.1	43.0			229
AVL: Econ	High	40.3	39.8	8.7	40.2	—	11.37**	228
	Low	42.3	43.1	7.4	42.5			216
AVL: Aesth	High	36.6	38.9	10.1	38.3	+	13.18**	228
	Low	34.4	35.0	7.2	35.7			216

TABLE 2—Continued

Test variable	Aptitude group	M _x ^a	M _y ^b	SD _y	My adj ^c	Δ cov ^d	F _M ^e	N
AVL: Soc	High	34.6	35.5	7.1	35.9	—	4.63*	228
	Low	36.2	37.5	6.2	37.1			216
AVL: Polit	High	42.8	42.5	7.2	42.6	—	.95	228
	Low	43.0	43.2	6.2	43.1			216
AVL: Relig	High	38.3	36.6	10.6	36.8	—	7.88**	228
	Low	39.0	39.1	8.5	38.9			216
D Scale	High	152.8	143.8	27.8	147.8	—	19.89**	228
	Low	172.6	162.7	27.3	158.7			224

^a Initial mean test score.

^b Retest mean score.

^c Retest mean adjusted by covariance procedures for initial test score.

^d Direction of change in covariance analysis; + = high-aptitude group had higher My adj., and — = high-aptitude group had lower My adj. compared with that of low aptitude group.

^e *F* ratio obtained in testing for homogeneity of adjusted means.

* *p* = .05.

** *p* = .01.

presented in Tables 2, 3, and 4. Consideration of the entries in the columns M_x (mean of initial test score) and M_y (mean of final test score) yields several observations relative to procedural matters. First, there were apparent differences between means on many variables for males and females. Therefore, the separate treatment of the two sexes appears sound. Second, there were apparent differences between initial means on many variables for high- and low-aptitude groups. This is as expected in view of the many reports of modest but significant correlations between aptitude and non-intellectual variables among Ss of the type employed, and makes the decision to use covariance procedures in the study appear sound. Third, it is apparent that there often were substantial differences between initial level of performance and that of 2 or 4 years later. Whereas these considera-

tions are pertinent to the design of the study, they are not the object of the present investigation. Consequently, no statistical evidence for these considerations is offered here. The interested reader is referred to earlier studies cited where these dimensions have been of direct interest.

Comparison of the Two Aptitude Groups

Final means, when adjusted by analysis of covariance procedures for initial performance, appear in Tables 2, 3, and 4 under the column headed My_{adj}. For convenience, the adjacent column, headed Δ_{cov}, shows the relationship between the level of adjusted means and level of aptitude; a + sign indicates that the high-aptitude group had the higher adjusted mean. The statistical tests for the covariance analysis appear in the columns headed F_M. These give the *F* ratios appropriate to

TABLE 3

COMPARISONS OF PERSONALITY TEST SCORE CHANGES OVER TWO YEARS FOR
HIGH- AND LOW-APTITUDE FEMALES: STUDIES 1, 2, AND 3

Test variable	Aptitude group	Mx	My	SDy	My adj	Δ cov	F_M	N
Study 1 E Scale	High	76.7	69.0	22.7	75.5	—	1.24	100
	Low	99.3	85.5	26.1	79.0			100
Study 2 D Scale	High	147.5	136.2	23.9	140.2	—	.45	210
	Low	161.5	145.8	28.3	141.7			210
E Scale	High	69.3	62.7	17.4	68.9	—	7.12**	210
	Low	90.3	79.6	22.7	73.4			210
F Scale	High	111.2	97.9	21.7	103.6	—	2.21	210
	Low	128.5	112.3	27.7	106.6			210
Study 3 CPI: Sy	High	25.1	25.3	5.4	24.2	+	.29	188
	Low	21.9	22.8	5.6	23.9			185
CPI: Sc	High	27.5	28.4	7.5	28.5	—	2.30	188
	Low	28.1	29.6	8.3	29.5			185
CPI: Ai	High	20.5	22.2	3.5	20.9	+	29.08**	188
	Low	15.8	17.6	3.8	19.0			185
CPI: Ie	High	38.6	40.6	4.9	39.0	+	21.08**	188
	Low	32.7	35.0	5.9	36.6			185
CPI: Re	High	32.5	33.0	4.0	32.1	+	2.82	188
	Low	29.2	30.4	4.8	31.4			185
AVL: Theor	High	39.3	40.2	7.4	40.1	+	8.83**	186
	Low	38.9	38.1	6.5	38.3			173
AVL: Econ	High	36.7	35.8	7.5	36.4	—	20.74**	186
	Low	39.4	40.2	6.8	39.6			173
AVL: Aesth	High	41.5	43.5	9.6	42.1	+	10.11**	186
	Low	37.1	38.0	8.6	39.5			173
AVL: Soc	High	39.5	40.7	6.9	41.2	+	.36	186
	Low	41.7	41.4	6.7	40.8			173

TABLE 3—Continued

Test variable	Aptitude group	M_x	M_y	SD_y	$M_y \text{ adj}$	$\Delta \text{ cov}$	F_M	N
AVL: Polit	High	37.9	37.1	7.4	37.4	—	6.20*	186
	Low	39.3	39.3	5.7	38.9			173
AVL: Relig	High	45.0	42.5	9.6	42.0	—	2.19	186
	Low	43.2	42.6	7.9	43.1			173
D Scale	High	148.1	139.7	26.8	144.0	—	9.81**	187
	Low	165.4	155.9	26.1	151.7			187

* $p = .05$.** $p = .01$.

testing homogeneity of adjusted means.²

Differences between adjusted means proved to be significant in 19 out of 40 possible instances; 10 of these were with male Ss and nine with female Ss. In the tables, significance at the .05 level is indicated by an asterisk following the F ratio; a double asterisk indicates significance at the .01 level. For males, the following differences were significant: F_2 (.01), CPI variables A_i (.01), I_e (.01), and R_e (.05), AVL variables T (.01), E (.01), A (.01), S (.05), and R (.01), and D_3 (.01). For females, the following dif-

ferences were significant: E_2 (.01), CPI variables A_i (.01) and I_e (.01), AVL variables T (.01), E (.01), A (.01), and P (.05), D_3 (.01), and E_5 (.05).

Differences significant at the .05 level number only four for the entire study. It is probably best to give them little attention since one or more could readily occur by chance, and since the actual difference in adjusted means in each case is less than one-quarter of the standard deviation of the variable in question.

Direction of Change in Relation to Aptitude Level

The direction of the outcome of each test is most easily studied by noting entries in the column headed Δ_{cov} . It will be remembered that "+" indicates a higher adjusted mean score for the high-aptitude group, and therefore more relative "gain" in score on the nonintellectual variable for more apt Ss. It appears that the direction of the difference is related to the general tendency for the score on the nonintellectual variable to increase or decrease from initial test to retest. To form a rough index, one may sum the initial means for the two aptitude

² Since homogeneity of slope of regression of final score on initial score is a precondition to the test of significance between adjusted means, F ratios appropriate to testing homogeneity of slopes were calculated. For males, one out of 20 F ratios (AVL S: Study 3) was significant at the .05 level as was one of 20 for females (AVL T: Study 3), and one for females was significant at the .01 level (F Scale: Study 5). In view of the 40 opportunities for significance to be discovered, and in view of the lack of consistency from study to study, it is not unreasonable to conclude that the significant instances were false positives attributable to chance. On these grounds, it was deemed appropriate to consider the evidence of all F ratios conducted on adjusted means.

TABLE 4

COMPARISONS OF PERSONALITY TEST SCORE CHANGES OVER FOUR YEARS FOR HIGH- AND LOW-APTITUDE MALES AND FEMALES: STUDIES 4 AND 5

Test variable	Aptitude group	Mx	My	SDy	My adj	Δ cov	F_M	N
Males Study 4 E Scale	High	78.6	64.7	17.5	69.1	—	2.26	34
	Low	98.0	80.6	22.9	76.1			34
Study 5 D Scale	High	153.6	140.2	26.9	143.9	+	.31	112
	Low	165.9	146.0	24.7	142.2			111
E Scale	High	80.7	69.2	23.6	73.8	—	.95	112
	Low	94.8	81.2	25.6	76.6			111
F Scale	High	116.7	98.2	23.7	103.5	—	.53	112
	Low	131.9	110.9	23.6	105.5			110
Females Study 4 E Scale	High	77.0	59.6	24.1	66.5	+	.39	35
	Low	96.1	71.4	28.7	64.3			34
Study 5 D Scale	High	148.0	132.1	27.4	137.2	—	.39	146
	Low	164.1	144.0	29.3	138.9			146
E Scale	High	69.6	58.6	19.4	65.3	—	5.45*	146
	Low	89.8	77.2	24.9	70.4			146
F Scale	High	111.8	92.4	22.8	99.2	—	3.39	146
	Low	130.4	110.9	30.7	104.1			146

* $p = .05$.

groups and do likewise for the retest means. If the question is then raised as to the number of instances in which a "general" increase in score over time is coupled with the occurrence of a higher adjusted mean score for the high-aptitude group and in which a "general" decrease is associated with a lower adjusted mean score for the high-aptitude group, it will be found that this condition exists in 33 out of the 40 possible instances in the study. The seven exceptions are: the CPI variable Sc for both males and fe-

males, the AVL variables T, E, and S for males, E₄ for females, and D₅ for males. Of the exceptions, only the last two are characterized by substantial change over time. If this analysis is reduced to consideration of the 15 instances in which a difference between aptitude groups was found to be significant at the .01 level, it will be observed that in 13 cases this same relationship prevails. In the remaining two (AVL variables T and E for males), the general shift in level of scoring over time is minimal. In other

words, when there is an overall change, there is a substantial tendency for the high-aptitude group to show it more strongly.

Change as a "Psychologically Positive" Phenomenon

There seems to be general agreement among psychologists that lower D, E, or F Scale scores indicate more "psychologically positive" behavior than do higher scores. For these variables, the aptitude group with the lower adjusted My score was designated as the group with the more "psychologically positive" behavior.

Either explicit or implicit in the numerous validity studies reported by Gough (1957) for the CPI scales is the notion that higher Sy, Sc, Ai, Ie, or Re scale scores indicate more "psychologically positive" behavior than do lower scores. For the CPI variables, the aptitude group with the higher adjusted My score was designated as the group with the more "psychologically positive" behavior.

It is difficult to determine which scales of the AVL are reasonably subject to this kind of designation. There does seem to be agreement among research psychologists that higher scores on the AVL Theoretical (T) and Aesthetic (A) scales are more "valued" in samples of bright and well-educated Ss than are lower scores. This value judgment is shared by the present writers, also. Therefore, for the T and A scales of the AVL, the aptitude group with the higher adjusted My score was designated as the group with the more "psychologically positive" behavior. No decision was made for the E, S, P, or R scales of the AVL.

Table 5 summarizes the results of the covariance analyses in a manner designed to facilitate study of the relationship between differences in adjusted performance level of the two aptitude groups on each variable and

the characteristic of "psychological positiveness." The question under consideration is whether high-aptitude groups have the more "psychologically positive" nonintellectual scale scores at the end of 2 or 4 years.

Consideration of the table shows that 32 of the 40 comparisons have a designation of more "psychologically positive" score, in accord with the designations specified in the paragraphs immediately preceding. For males, 14 of the 16 available comparisons favor the high-aptitude group. Of the 16 comparisons, seven showed a significant difference between high- and low-aptitude groups, and in all seven the high-aptitude group is favored. For females, the same figures are descriptive of the results. Fourteen of the 16 available comparisons favor the high-aptitude group, and of the 16 comparisons, the seven showing a significant difference between high- and low-aptitude groups all favor the high-aptitude group.

There is, therefore, strong evidence that high aptitude is associated with growth in a "psychologically positive" direction. The trend in development of nonintellectual characteristics appears to be similar to that of the development of intellectual characteristics in that for brighter young people greater development in a desirable direction takes place over time.

Comparison of the Sexes

Examination of the 20 parallel comparisons (same study and variable) between males and females shows that the direction of difference between high- and low-aptitude groups (Δ_{cov}) was the same in 17 instances. The exceptions were the AVL variable S, and E₄ and D₅. It may be noted that none of the 15 differences significant at the .01 level was involved in these exceptions.

Of the 16 parallel comparisons in

TABLE 5
SUMMARY OF COVARIANCE ANALYSES AND "MOST PSYCHOLOGICALLY POSITIVE"
TEST SCORES FOR ALL GROUPS AND ALL STUDIES

Variable	Males		Females	
	Aptitude group with highest My adj.	Aptitude group with most "psychologically positive" My adj.	Aptitude group with highest My adj.	Aptitude group with most "psychologically positive" My adj.
Study 1				
E	Low	High	Low	High
Study 2				
D	Low	High	Low	High
E	Low	High	Low**	High
F	Low**	High	Low	High
Study 3				
CPI Sy	High	High	High	High
CPI Sc	Low	Low	Low	Low
CPI Ai	High**	High	High**	High
CPI Ie	High**	High	High**	High
CPI Re	High*	High	High	High
AVL Theor	High**	High	High**	High
AVL Econ	Low**	?	Low**	?
AVL Aesth	High**	High	High**	High
AVL Soc	Low*	?	High	?
AVL Polit	Low	?	Low*	?
AVL Relig	Low**	?	Low	?
D	Low**	High	Low	High
Study 4				
E	Low	High	High	Low
Study 5				
D	High	Low	Low	High
E	Low	High	Low*	High
F	Low	High	Low	High

* $p = .05$.

** $p = .01$.

which it seemed possible to designate the direction of psychologically desirable development, 14 were instances in which the aptitude group having the more psychologically positive adjusted mean was the same for males and females. The exceptions were E_4 and D_5 .

SUMMARY AND CONCLUSIONS

All investigators of the effects of college attendance upon student personality have reported measurable effects of a generally desired sort (e.g., reduction in prejudice). Local studies have demonstrated that similar personality changes occurred in young adults with no college attendance,

some college attendance, and with the maximum amount of college attendance possible in 2-year and 4-year periods.

In light of the empirical facts that college or college-bound Ss are brighter than average, and that intellectual changes continue well into adulthood for Ss who were initially brighter than average, a series of studies was undertaken to determine if the trend in nonintellectual development was similar to that of intellectual development for initially brighter-than-average Ss. If so, it was felt that this might account in large part for the positive findings reported in the "college impact" research. Existing data

from five separate longitudinal studies of college-bound young adults were utilized to determine whether or not nonintellectual changes occurred differentially for groups of differing academic aptitude. Measures involved were the Ethnocentrism Scale, the Dogmatism Scale, the Authoritarianism Scale, five scales from the California Psychological Inventory, and the six scales of the Allport-Vernon-Lindzey Study of Values.

Considering that separate studies were conducted for males and females, 40 comparisons of nonintellectual change over time for high- and low-aptitude groups were available. Re-test score means were adjusted by covariance procedures for initial test score, and significance tests were performed. Of the 40 comparisons, 15 were significant at the .01 level, and four were significant at the .05 level.

In 33 of 40 possible instances, an overall increase in score over time was coupled with evidence of greater increase for the high-aptitude groups, or an overall decrease in score was coupled with greater relative decrease for the high-aptitude groups. It appeared possible to make a judgment as to which of a pair of adjusted means in a given comparison represented the more "psychologically positive" development for 32 of the 40 comparisons. Twenty-eight of these favored the high-aptitude group.

The following conclusions appear justified:

1. There is substantial evidence to suggest that there are differences between more and less apt young adults in the relative degree of change in nonintellectual scale scores over time. The findings exhibit a substantial tendency for young adults of higher aptitude to exhibit more nonintellectual change over time and in the direction of the trend of college students in general.

2. There is strong reason to believe that nonintellectual changes which occur over time among young adults are generally of a more "psychologically positive" nature as defined by values of our culture or of the subculture of educated persons.

3. There is strong evidence that high-aptitude groups of young adults have substantially more "psychologically positive" nonintellectual development over time.

4. The evidence indicates that nonintellectual development is similar in its trend to that of intellectual development for initially brighter-than-average young adults.

5. The several results pertaining to nonintellectual development observed in this study appear to be similar for males and females.

6. The results pertaining to nonintellectual development appear to be of some generality in regard to different samples of young adults and different kinds of nonintellectual measures.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of values: Manual*. (3rd ed.) Boston: Houghton Mifflin, 1960.
- BAYLEY, N. On the growth of intelligence. *American Psychologist*, 1955, **10**, 805-818.
- BAYLEY, N. Data on the growth of intelligence between 16 and 21 years as measured by the Wechsler-Bellevue Scale. *Journal of Genetic Psychology*, 1957, **90**, 3-15.
- BAYLEY, N., & ODEN, M. H. The maintenance of intellectual ability in gifted adults. *Journal of Gerontology*, 1955, **10**, 91-107.
- BLOOM, B. S., & WEBSTER, H. The outcomes of college. *Review of Educational Research*, 1960, **30**, 321-333.
- CONRY, R., & PLANT, W. T. WAIS and group test predictions of an academic success criterion: High school and college. *Educational and Psychological Measurement*, 1965, **25**, 493-500.
- COREY, S. M. Attitude differences between

- college classes: A summary and criticism. *Journal of Educational Psychology*, 1936, **27**, 321-330.
- COREY, S. M. Changes in the opinions of female students after one year at a university. *Journal of Social Psychology*, 1940, **11**, 341-351.
- CRONBACH, L. J. *Essentials of psychological testing*. (2nd ed.) New York: Harper, 1960.
- FOSTER, J., STANEK, R., & KRASSOWSKI, W. *The impact of a value-oriented university on student attitudes and thinking*. Final Report, 1961, Cooperative Research Project No. 729, University of Santa Clara, California.
- FREEDMAN, M. B. *Impact of college*. New Dimensions in Higher Education, No. 4, Washington, D. C.: U. S. Department Health, Education & Welfare, Office of Education, 1960.
- GOUGH, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *Journal of Social Psychology*, 1951, **33**, 237-246.
- GOUGH, H. G. *California psychological inventory: Manual*. Palo Alto, California: Consulting Psychologists Press, 1957.
- JACOB, P. *Changing values in college*. New York: Harper, 1957.
- KUHLEN, R. G. Changes in the attitudes of students and relations of test responses to judgments of associates. *School and Society*, 1941, **53**, 514-519.
- LEHMANN, I. J. Changes in critical thinking, attitudes, and values from freshman to senior years. *Journal of Educational Psychology*, 1963, **54**, 305-315.
- LEHMANN, I. J., & DRESSEL, P. L. *Critical thinking, attitudes, and values in higher education*. Final Report, 1962, Cooperative Research Project No. 590, Michigan State University.
- LEHMANN, I. J., & DRESSEL, P. L. *Changes in critical thinking, attitudes, and values associated with college attendance*. Final Report, 1963, Cooperative Research Project No. 1646, Michigan State University.
- LEHMANN, I. J., SINHA, B. K., & HARTNETT, R. T. Changes in attitudes and values associated with college attendance. *Journal of Educational Psychology*, 1966, **57**, 89-98.
- MCCULLERS, J. C., & PLANT, W. T. Personality and social development: Cultural influences. *Review of Educational Research*, 1964, **34**, 599-610.
- NEWCOMB, T. M. *Personality and social change*. New York: Dryden Press, 1943.
- OWENS, W. A., JR. Age and mental abilities: A longitudinal study. *Genetic Psychology Monographs*, 1953, **48**, 3-54.
- PLANT, W. T. Changes in ethnocentrism associated with a four-year college education. *Journal of Educational Psychology*, 1958, **49**, 162-165. (a)
- PLANT, W. T. Changes in ethnocentrism associated with a two-year college experience. *Journal of Genetic Psychology*, 1958, **92**, 189-197. (b)
- PLANT, W. T. *Personality changes associated with a college education*. Final Report, 1962, Cooperative Research Project No. 348, San Jose State College, California.
- PLANT, W. T. Longitudinal changes in intolerance and authoritarianism for subjects differing in amount of college education over four years. *Genetic Psychology Monographs*, 1965, **72**, 247-287.
- PLANT, W. T., & LYND, C. A validity study and a college freshman norm group for the Wechsler Adult Intelligence Scale. *Personnel and Guidance Journal*, 1959, **37**, 578-580.
- PLANT, W. T., & RICHARDSON, H. The I.Q. of the average college student. *Journal of Counseling Psychology*, 1958, **5**, 229-231.
- ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychological Monographs*, 1956, **70**(18, Whole No. 425).
- TELFORD, C. W., & PLANT, W. T. *The psychological impact of the public two-year college on certain non-intellectual functions*. Final Report, 1963, Cooperative Research Project No. 914, San Jose State College, California.
- WEBSTER, H. Changes in attitudes during college. *Journal of Educational Psychology*, 1958, **49**, 109-117.
- WEBSTER, H., FREEDMAN, M. B., & HEIST, P. Personality changes in college students. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1962. Pp. 811-846.

(Received May 23, 1966)

LEARNER COMPETENCE, MODEL COMPETENCE, AND NUMBER OF OBSERVATION TRIALS IN VICARIOUS LEARNING¹

FREDERICK H. KANFER AND PRYSE H. DUERFELDT

University of Oregon Medical School

The effects of vicarious learning were studied in a paired associate nonsense syllable task under conditions of varied model competence, varied subject competence, and varied numbers of vicarious trials. 120 Ss heard a model who was at an early or a late stage of learning. The model was heard at either an early or a late stage in Ss' learning, and Ss heard the model for either 1 or 3 of 10 trial blocks. 15 control Ss learned only under direct reinforcement. Ss exposed to the model early in acquisition learned significantly better than Ss exposed late. Model competence and duration of exposure did not affect learning significantly. The results suggested that vicarious trials late in acquisition had a disruptive effect, while early vicarious exposure yielded benefits similar to those of direct reinforcement trials.

While observational learning has long been used as a teaching adjunct in educational practices, attention to the specific parameters of this vicarious learning process has been recent. Studies by Bandura (1962, 1965) and his coworkers have demonstrated the learning of novel responses and the inhibition and extinction of previously learned responses as a function of observation of a model. Kanfer and Marston (1963) have compared the effectiveness of vicarious and direct reinforcement and have shown that the percentage of reinforcement affects the observer's performance (Marston & Kanfer, 1963).

Since a subject's (S's) disposition to attend to the model's behavior is crucial to his reception and utilization of information, it is apparent that the relationship between an S and a model represents an important variable in vicarious learning (Bandura, 1965). One aspect of this relationship, the

apparent competence of the model, was examined by Rosenbaum and Tucker (1962). They found that greater apparent competence of the model enhanced learning in the observer. These results were attributed to the fact that S's prior social history affects his use of the information provided by the model, that is, S has learned that more competent models provide more efficient and reliable information to the observer. A similar conclusion was reached by Bandura and Kupers (1964) in a study comparing the effectiveness of adult and peer models on imitated behavior in children.

From the learner's viewpoint, his own competence level relative to the model would further modify the relationship. In observing a relatively incompetent model, S would be expected to use the available intertrial time for rehearsal of his already acquired responses rather than attending to the model's performance for additional information. In fact, continued attention to a model who makes predominantly incorrect responses should result in an interference effect on S.

¹This study was supported in part by research grant MH 06922-04, from the National Institutes of Mental Health, United States Public Health Service, to the first author.

The present study was designed to consider these several competence parameters of the model-subject relationship with competence defined not by model status or experimenter instructions, but by the model's actual performance. In a three-factor design the following variables were studied with a verbal paired associate learning task: (a) the number of observation trials, (b) the acquisition stage (competence) of the learner, and (c) the acquisition stage of the model.

METHOD

Preparation of Model Tapes

The learning task consisted of a list of 10 pairs of nonsense syllables of mean association value. The list was taken from Hovland's (1939) list. Pairs which showed the greatest variability in being learned or not learned and pairs in which one member was similar in pronunciation to that of another pair were discarded. This selection yielded a list of 10 paired associate nonsense syllables which constituted *S*'s learning task. Ten random orders of presentation were repeated for 40 trials. There was a 3-second interval between presentation of syllable pairs, a 3-second interval between presentation of the stimulus member and response member of each pair, and a 2-second exposure of the pair. At the end of each list of 10 pairs a 4-second intertrial interval was given. Five presentations of the list constituted one block of trials.

Errors on the taped presentations were taken from actual performances obtained from pilot *Ss*. Thus, both intrusions and partially correct syllables appeared. For the model at an *early* learning stage, the tapes contained the following number of correct responses on each trial: Trial 6:1; 7:1; 8:2; 9:3; 10:3. On Trial 11:3; 12:3; 13:3; 14:4; 15:4. On Trial 16:5; 17:5; 18:5; 19:6; 20:6. For the tapes representing a model at a *late* learning stage, the following number of correct responses were used on each trial: 21:6; 22:6; 23:6; 24:6; 25:7. On Trial 26:7; 27:7; 28:8; 29:8; 30:8. On Trial 31:8; 32:8; 33:8; 34:8; 35:8. The tapes represented acquisition curves with 1-6 correct responses per block for the model at the early acquisition stage, and 6-8 correct responses for the model representing a later learning stage.

The tapes were prepared by the first au-

thor and a graduate assistant. They represented two variables of the design, the stage of model learning, and the number of observational trials. The entire learning task was taped so that each group heard the stimulus word, an interval for *S*'s response, and the stimulus-response pair on each trial. On those blocks on which vicarious experience was offered, *S* also heard a model response between the stimulus and the stimulus-response presentations. Following each set of 10 pairs, the reader on the tape said, "New trial."

Experimental Design

A control group learned the list without vicarious trials. The following groups constituted a $2 \times 2 \times 2$ factorial design. In the group designation the numbers refer to the blocks on which vicarious trials were given and the letters *E* or *L* refer to an *early* or *late* model stage of learning. For one-block observation, *Ss* were exposed to the model's performance on Block 3, either with a model simulating learning during the third block of acquisition (3-*E*), or at a late stage, Block 6 of acquisition (3-*L*). Two additional groups were given the one-block vicarious experience during Block 6, either with an early model (6-*E*), or with a late model (6-*L*).

For groups with three blocks of observation, the same pattern was replicated. Two groups heard either an early model tape, simulating Blocks 2, 3, and 4 of acquisition (2, 3, 4-*E*), or a late model simulating Blocks 5, 6, and 7 of acquisition (2, 3, 4-*L*). Two groups obtained their vicarious experience on Trials 5, 6, and 7, either with an early model (5, 6, 7-*E*), or with a late model (5, 6, 7-*L*).

Subjects

All *Ss* were volunteer college students who were paid for participation in this study. One hundred and thirty-five *Ss* were run, with random assignment of 15 *Ss* per group.

Procedure

The *S* was comfortably seated in an experimental room facing a partition behind which the experimenter (*E*) operated a tape recorder and recorded *S*'s responses. Prior to *S*'s arrival, *E* selected the appropriate tape. The instructions were taken in part from Hovland (1939), modified to fit the present study. All *Ss* were told that this was an experiment in learning nonsense syllables. They would hear pairs of nonsense

syllables over a tape recorder and their job was to learn the pairs so that upon hearing the first syllable, they would be able to respond immediately with the second. All Ss, except those in the control group, were also told:

People find it easier to learn this kind of task when there is a break. Sometime at the end of one of the trials I will interrupt and tell you that it is time for the break. Since many people rehearse during this break anyway, we have decided to let you listen to a subject who took part in this experiment the other day, so that everybody gets some benefit from the break. Then I will tell you when it is time to begin again.

The Ss were encouraged to speak clearly and told that it was a very difficult task.

At the beginning of a vicarious learning section, Ss were told: "All right, it's time for the break. Listen, but don't respond. I will tell you when the break is over." At the end of the vicarious learning section they were told to respond again. This experimental design thus yielded learning data only for those blocks on which S was required to respond.

RESULTS

All S responses were recorded in blocks of five list presentations. An analysis of variance was run on the total number of correct responses over all blocks in each group in order to determine the effect of sex differences. None of the analyses revealed a significant difference between male and female students in each of the groups. Therefore, data from both sexes were pooled for subsequent analyses.

Since Ss in various groups were exposed to either one or three blocks of vicarious learning during which no responses were given, and since these periods occurred at different stages of learning, the main effects of the treatment variables could best be compared by an analysis of the final acquisition block (Block 8). A factorial analysis was carried out for exposure (one block versus three blocks), learner stage (early versus late); and model stage (early versus late).

The analysis of variance indicated that duration of exposure to the taped model did not affect learning. There was also no difference between Ss who listened to an early model and those who listened to a late model who was more competent on this task. The only significant difference ($F = 4.38, p < .05$) was found between groups in which Ss were exposed early during acquisition and those who were exposed near the end of acquisition training (learner stage). The Ss who heard the model early gave more correct responses than Ss who had the same experiences at a later learning stage. The combined early stage groups yielded a mean of 44.4 correct responses on Block 8, while Ss who were in groups with vicarious learning during later acquisition blocks gave a mean of 40.4 correct responses. A comparison with the control group average of 43.7 correct responses for Block 8 suggests that the obtained difference was due primarily to poor learning in groups which were exposed to a model tape late in acquisition. This finding suggests that vicarious learning may be more effective in earlier stages than in later stages.

An alternate explanation of these findings might be that the vicarious experience had a disruptive effect and Ss in the late groups were exposed to this disruption after more learning and closer to the test for acquisition in Block 8. The disruption hypothesis was tested by comparing the performance of the experimental groups on acquisition blocks immediately following their vicarious learning exposure with the performance of the appropriate nonexposed control groups. An analysis of variance was carried out on Block 4 comparing Groups 3-E and 3-L with 6-E and 6-L. While the former groups had experienced vicarious learning on a preceding block (Block 3) the later group served

as controls without previous vicarious experience. The analysis revealed no significant difference between these groups. A further test was carried out by an analysis of variance on Block 5 for Groups 2, 3, 4-E and 2, 3, 4-L versus 6-E and 6-L. This analysis also failed to reveal significant differences between the groups. For Ss who received the vicarious experience early in acquisition, either for one or three blocks, there was no apparent disruption effect and their performance was equal to that of Ss who were never exposed to vicarious learning up to that stage.

Comparisons for each of the trial blocks were not possible since all groups yielded data only on Blocks 1 and 8. However, a further test of the effects of the timing of vicarious learning was made by comparing 3-E and 3-L with 6-E and 6-L on Block 7. A factorial analysis of variance was carried out for these four groups with learner stage and model stage as the main factors. The analysis yielded a significant F -ratio (4.12, $p < .05$) only for the learner stage. The Ss who heard the model on Block 3 yielded more correct responses than Ss who had the vicarious experience on the preceding (sixth) block. These supplementary analyses suggest that exposure to models had a disruptive effect only during late acquisition trials. Thus, the disruptive effect was not simply a function of the interval between the vicarious experience and the block on which the performance was analyzed. The results tend to confirm the hypothesis that a disruptive effect occurs mainly when Ss are near mastery of the learning task. These results seem to be unaffected by the relative competence of the taped model.

DISCUSSION

The results of this study confirm earlier observations (Bandura, 1965; Kanfer, 1965) that knowledge of the specific parameters of vicarious learning is an essential step toward its refinement as a practical educational tool. The finding that Ss derive more benefit from observational learning during the early, rather than the later, stages of their attempts to master a task clearly suggests that the efficient use of such techniques as motion pictures or demonstrations as training aids is dependent upon their time of presentation. The correlated finding that an S near mastery of a task is hindered by observing a model at or below his level of competence suggests that pursuit of such a procedure as an educational adjunct may discourage student and teacher alike.

The fact that no differences were found between the direct learning and either of the early observational learning groups supports past findings (Berger, 1961; Kanfer & Marston, 1963) that observation of a model can effectively *replace* much of the trial and error learning typically utilized in developing task mastery. In conjunction with the finding that the amount of direct learning replaced by observation did not differentially affect performance, a tentative practical implication is suggested: When observational learning is more economical in terms of time, money, danger, etc., than direct learning, it may effectively replace large blocks of the trial and error method and promote the mastering of a task with equal efficiency.

Unlike other studies in which model competence has proven an important variable (Bandura & Kupers, 1964; Rosenbaum & Tucker, 1962), "competence" was here defined by the

accuracy of the model's performance (as determined by E), not by his status or implied response accuracy (as appraised by S). Unless S views the model as highly competent or highly incompetent it appears unlikely that he attends differentially to the models at different competence levels, regardless of their accuracy. This post hoc explanation of why model competency was not an effective variable in the present study is given some support in a recent unpublished dissertation by Craig (1964). It was found that with model accuracy held constant, S s requested assistance from models labeled as highly competent significantly more than from models labeled as low in competence. An important factor which may influence the contributions of observed trials to learning is S 's set as controlled by instructions. The present study maximized the effects of the vicarious trials by implying the potential benefits of utilizing the interval designated as a "break." Together with the fairly high motivation in the volunteer college student population sampled here, these variables would tend to reduce differences in model competence and to maximize vicarious learning. The present findings must therefore be limited in generalization to similarly favorable instructional conditions. The study demonstrates that observational learning can be an effective replacement for much of traditional trial and error learning when the knowledge of its parameters are taken into consideration.

REFERENCES

- BANDURA, A. Social learning through imitation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1962*. Lincoln: University of Nebraska Press, 1962. Pp. 211-269.
- BANDURA, A. Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. II. New York: Academic Press, 1965. Pp. 1-55.
- BANDURA, A., & KUPERS, C. J. Transmission of patterns of self-reinforcement through modeling. *Journal of Abnormal and Social Psychology*, 1964, **69**, 1-9.
- BERGER, S. M. Incidental learning through vicarious reinforcement. *Psychological Reports*, 1961, **9**, 477-491.
- CRAIG, K. D. Partner competency and prior success-failure experiences as determinants of manding in dyadic interactions. Unpublished doctoral dissertation, Purdue University, 1964.
- HOVLAND, C. I. Experimental studies in rote learning theory. IV. Comparison of reminiscence in serial and paired associate learning. *Journal of Experimental Psychology*, 1939, **24**, 466-484.
- KANFER, F. H. Vicarious human reinforcement: A glimpse into the black box. In L. Krasner & L. P. Ullmann (Eds.), *Research in behavior modification: New developments and their clinical implications*. New York: Holt, Rinehart and Winston, 1965. Pp. 244-267.
- KANFER, F. H., & MARSTON, A. R. Human reinforcement: Vicarious and direct. *Journal of Experimental Psychology*, 1963, **65**, 292-296.
- MARSTON, A. R., & KANFER, F. H. Group size and number of vicarious reinforcements in verbal learning. *Journal of Experimental Psychology*, 1963, **65**, 593-596.
- ROSENBAUM, M. E., & TUCKER, I. F. The competence of the model and the learning of imitation and nonimitation. *Journal of Experimental Psychology*, 1962, **63**, 183-190.

(Received June 8, 1966)

EFFECTS OF TRAINING PROCEDURES ON CREATIVITY TEST SCORES

DENNIS R. RIDLEY

AND

ROBERT C. BIRNEY

University of California, Santa Barbara

Amherst College

159 college freshman males participated in experiments concerned with the effect on original behavior, as measured by 2 tests from Guilford's originality test battery (Unusual Uses and Plot Titles), of the following variables: (a) instructions to be original on the Unusual Uses tests; (b) word-association training; and (c) training in the use of problem-solving heuristics for the Unusual Uses test. Major hypotheses, based on previous literature, were that all variables would significantly facilitate performance on the Unusual Uses test, and that the 2nd variable would significantly facilitate performance on the Plot Titles test. Although these hypotheses were confirmed, a further hypothesis dealing with word-association training, derived by Maltzman (1960) from S-R principles, was not supported. The results suggest a testable approach to operational training of originality.

This study reports two experiments in which an attempt was made to enhance creative performance. Those tests developed by Guilford (1950, 1957) which call for divergence of response to relatively simple tasks were used. Maltzman (1960), in a review of the literature, points out that early attempts to enhance creativity (e.g., Osborn, 1957) were nonsystematic in nature. *Defining originality as statistically infrequent but relevant behavior*, he has reported a series of experimental attempts to produce original behavior by using a modified free-association technique (Maltzman, Simon, Raskin, & Licht, 1960). Increases in scores on the Guilford Unusual Uses (UUT) and Word Association (WAT) tests were obtained with prior repetition of word association in these experiments, but Gallup (1963) subsequently obtained similar increases with prior arithmetic and vocabulary rehearsal, casting doubt on the assumption that the training needed to be task-relevant. Rather the possibility that subjects (Ss) changed sets from speed to originality is raised. Early work by the Guilford group had already suggested this effect (Christensen, Guilford, & Wilson,

1957). Rosenbaum, Arenson, and Panman (1964) used both instructions to make "original" word association and five repetitions of word association, and found a training effect but no instructional effect on total productivity which produced increased originality as a secondary effect.

The experiments reported here test the major hypothesis that task-relevant principles exist for simple and complex tasks requiring divergence of response, that such principles can be deduced and expressed, that the principles can be taught, and that Ss who master these strategies will outperform those who have not. Following Maltzman et al. (1960) and Simon (1961), we also set out to replicate the word-association training effect, and to explore the effects of repetition well beyond any they attempted with length of word list held constant. Given their findings it was hypothesized that repetition of association is an effective means of increasing scores on the Guilford UUT and Plot Titles Test (PTT). The influence of "set" is chiefly determined by the instructions used, and to maximize such effects we introduced an instructional variable whereby the experimental

groups were given written instructions to perform as creatively as possible. By using two-way analysis of variance designs for each experimental condition we were able to test for the influence of instructions, word-association repetition, and training in the "heuristics" of divergent thinking on the UUT and the PTT. It was expected that all three variables would increase UUT scores, and that word association would produce significant carry-over to PTT scores; thus leaving the effects of instructions and heuristics training on the PTT matters for exploration. The design in Table 1 shows that two distinct experiments were conducted, each testing for the effects of the instructional variable as it interacts with a different type of training.

One persistent artifact in creativity research is that increasing the quantity of responses from Ss may produce increased frequency of originality. Christensen et al. (1957) have shown that this need not be so, but we present additional analysis for PTT data to assess the effects of increased levels of response.

METHOD

Subjects

Amherst freshmen ($N = 159$) were randomly selected from the Amherst freshman picture book, randomly assigned to experimental or control conditions, and contacted individually. The contacting procedure was standardized, and if an individual could not come at his preassigned time, he was re-assigned to another time. Over 95% of those contacted participated.

Dependent Variables

In the UUT S is asked to write unusual uses for given objects and not to write the same use more than once. A revision of the standard instructions to the UUT was made in order to set Ss toward originality of response. The scoring procedure for the UUT was kept similar to that of Maltzman and

TABLE 1
THE EXPERIMENTAL DESIGN, SHOWING SIZE
OF GROUPS TESTED UNDER
EACH CONDITION

Instructions	Training		
	None	Heuristics	Word association
Standard	C 1 ($N = 22$)	X 1 ($N = 19$)	X 3 ($N = 9$)
	C 3A ($N = 12$)		X 4 ($N = 14$)
Revised	C 2 ($N = 18$)	X 2 ($N = 28$)	X 5 ($N = 8$)
	C 3B ($N = 14$)		X 6 ($N = 15$)

Simon, by employing empirical norms collected by Maltzman et al. (1960), and counting responses which occurred not more than once in the normative population. A modification was that certain responses (but very few) were eliminated as not acceptable according to Guilford's Scoring Guide.

The PTT consists of two story plots, for which Ss are instructed to write as many appropriate titles as possible in the time allowed. Two scores were obtained: number of titles (fluency) and originality. Plot titles were randomly divided among seven judges who were instructed to sort titles into three categories of quality as defined by Guilford's Scoring Key.

Training Materials

Heuristics. A booklet of "strategies" of heuristic value in the UUT was written from suggestions of others and from the first author's own thinking, and revised through pretesting to include five strategies presumed to be the clearest and most useful. The booklet contains illustrations of the use of each strategy and blank spaces for application to three common objects. For example, one strategy is the following: "Transform the object: burn it, cut it, paint it, etc. What uses of your object do these transformations suggest?" An example is: a brick (with a hole in it) as a pencil holder. A practice test with three new objects is included.

Word association. The stimulus words were taken from those used by Simon (1961) and by Maltzman et al. (1960). A tape recorder was used in order to present the

stimulus words uniformly over the conditions.

Training

The experimental treatments were administered about 2 weeks after contact was made. The two groups scheduled to receive heuristics training for the UUT (X 1 and X 2) came first on successive evenings. Word-association training was given to four groups (X 3, X 4, X 5, and X 6) on two different evenings. As shown in Table 1 the only difference between X 1 and X 2, between X 3 and X 5, and between X 4 and X 6 was in the instruction variable.

Heuristics training. Heuristics training was conducted in a manner that was judged by two confederates to be practically identical in X 1 and X 2. The Ss were told that they would soon be given a test in which they were to write down unusual uses for certain objects, and that the following session would consist of the presentation of a group of strategies which are thought to be helpful in thinking of unusual uses for things. The booklet of strategies was distributed. For each strategy Ss were to read the statement and the examples, and take 2 or 3 minutes to practice using the strategy on their own. Afterwards the experimenter (E) called for examples of uses for each of the objects. The E tried to express no approval or disapproval. The training period was followed by a short practice test.

Next the booklets were collected and the UUT and PTT were distributed in an attached form. The UUT is given in two parts of 5 minutes each, and the PTT is given in two parts of 3 minutes each. The instructions were the standard written instructions.

After both tests Ss were asked to indicate by the appropriate number which strategies, if any, they consciously employed for each use they listed. Sheets with spaces exactly corresponding to spaces on the UUT and headed with the numbered strategies were used. These sheets conveniently led to a "strategy score" for each S consisting of a tally of the instances in which strategy employment and unique uses were indicated on corresponding spaces.

Word-association training. The procedure for word-association groups was the same except that length of training was either 12 or 16 trials. The Ss were informed that they would first take a word-association test. They were told that the tape recorder would read a list of six words a total of either 12 or 16 times. Their instructions were to asso-

ciate to each word by writing down in the appropriate space the first word that came to mind, and to respond after the first reading with a different word than the ones they had used before. Trials were separated only by a pause and the recorded number of the trial.

After this training, E explained that a new list of six words would be read, and they were again to write down the first word that came to mind. Subsequently, the UUT and PTT were administered exactly as they had been for heuristics groups.

Next the booklets of strategies were distributed with the explanation that these booklets had been presented to other students to help them think of unusual uses. The E read the booklet to Ss as they followed. Then the booklets were collected and Ss were asked to indicate strategy employment just as the heuristics-trained groups had done.

Control Groups

There were three control groups (C 1, C 2, and C 3), the first two of which were controls for all training effects on the two posttests, while the third controlled for effects of word-association training on the WAT.

The procedure for C 1 and C 2 was identical except that C 2 took the revised form of the UUT. The UUT and PTT were administered in the same manner as for the experimental groups. The heuristics strategies were next introduced exactly as for the word-association groups, and Ss were asked to indicate the strategies which they had employed.

C 3 was originally scheduled to take only the WAT. However, in order to enlarge the number of control Ss for the UUT, the UUT was subsequently administered, approximately half taking each form of the test.

RESULTS

For the analysis the experiment using heuristics training will be referred to as Experiment I and the one using word-association training will be referred to as Experiment II. It will be convenient to divide the presentation of results between the results for the UUT (including strategy data) and the results for the PTT.

Results for the UUT

The scoring procedure described above was used to find the total number of unique uses for each *S*, where uniqueness was defined as occurring not more than once in the norms of Maltzman et al. (1960). Since the frequency distributions were highly skewed, a square root transformation ($\sqrt{X - .5}$) was applied to the scores.

Results of the analysis of variance of UUT results appear in Table 2. Instructions were effective at the .05 level of confidence in both Experiment I and Experiment II. Word-association training also had a significant effect on scores ($p < .05$). Heuristics training proved to have the most highly significant effect ($p < .001$). Interaction was negligible in both experiments.

These results confirm our hypotheses for the UUT concerned respectively with heuristics training,¹ word-association training, and instruction. Remaining to be considered are the analysis of strategy scores and a sub-hypothesis suggesting that extended word-association training will heighten the effects on creative performance.

The first evidence of strategy employment was the proportions of

¹To explore the possible effects of extended heuristics training, five *Ss* were recruited from a psychology class at Amherst. This class had been administered a form of the UUT with new objects in a standard procedure. The *Ss* participated in a procedure aimed at improving their performance on a retest of the UUT in which they worked on their own through a longer form of the booklet of strategies. After 1 month, the standard UUT was administered by standard procedures. Scoring was standard on the posttest, but the pretest had to be scored from norms which were collected from the sample itself. The reliability coefficient for the two forms of the UUT was +.90. All *Ss* but one improved markedly, creating the impression that extended acquaintance with strategies can have marked facilitating effects.

TABLE 2

ANALYSIS OF VARIANCE ON UUT SCORES

	Source		
	<i>df</i>	<i>MS</i>	<i>F</i>
Experiment I			
Instructions (A)	1	.11	4.23*
Heuristics (B)	1	.77	29.62**
A × B	1	.01	—
Error	109	.026	
Experiment II			
Instructions (A)	1	.32	5.82*
Word Association (B)	2	.28	5.09*
A × B	2	.005	—
Error	106	.055	

Note.—This analysis was done by the method for unequal *Ns* presented by Walker and Lev, 1953, pp. 381-382.

* $p < .05$.

** $p < .001$.

unique uses which involved the use of strategies in each group, that is, the ratio of total strategy scores to total number of unique uses. These ratios ranged from .42 to .64. Second, it was found by a *t*-test between X 2 and C 2 in mean strategy scores, at a significance level of $p < .01$, that heuristics treatments were associated with more successful strategy employment than control treatments.

To show the results of the extension of Simon's (1961) study the percentage of control performance gained by each amount of word-association repetition is displayed in Figure 1. Estimates from Simon's data are compared with our findings for the 12 and 16 levels under standard instructions. Appropriate *t*-tests showed the following: Both levels differ significantly from the control level, but the difference between them is not significant; only the 16 level was significantly greater under revised instructions. Thus, we have no evidence here that extended experience of word association will increase UUT performance

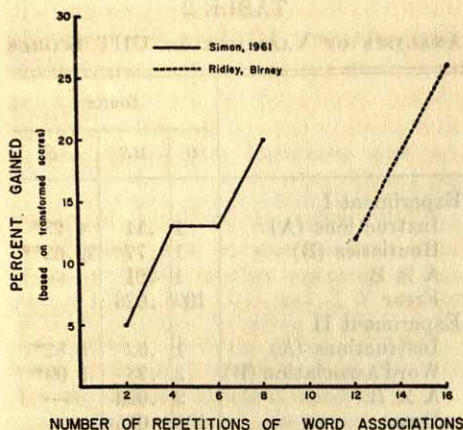


FIG. 1. Percentage of Control performance on UUT gained for each level of word-association rehearsal—no data for 10 and 14 repetitions. (Estimates from Simon's, 1961, data disregard length of word list.)

beyond that obtained with eight repetitions.

Results for the PTT

The scoring procedure presented above was used and only the responses to one of the two plots were scored. Reliability estimates for the seven judges were formed on the basis of scores for 40 titles which all judges scored in common. With one exception, these reliabilities fall between .50 and .75, which are comparable to reliabilities obtained by Wilson, Guilford, and Christensen (1953).

Since again the frequency distribution of scores were skewed, a square root transformation ($\sqrt{X + .5}$) was applied to all scores.

Analyses of variance for Experiments I and II on the PTT are summarized in Table 3. The instruction variable failed to show a significant effect in either experiment. There were significant effects of heuristics ($p < .05$) and word-association training ($p < .01$). Again, interaction was zero or negligible.

These data demonstrate the hypothesized effect of word-association training. Since our inspection of the other two variables for PTT analysis was regarded as exploratory, the heuristics effect and the lack of an instructional effect will require consideration later. We have yet to examine the relation between effects achieved on fluency and originality.

A *t*-test between the mean fluency of pooled control groups and of pooled experimental groups (excluding the 12 repetition row) disclosed that the latter had significantly higher production of titles ($p < .05$). The relevant question is whether the treatment influenced both fluency and originality or whether the higher originality was an artifact of the unexpected fluency effect. Product-moment correlations between fluency and originality for the pooled control groups and the four experimental groups here considered were high (+.90 and +.76, respectively). Although the drop in correlation is significant, suggesting a training effect of separating the two

TABLE 3
ANALYSIS OF VARIANCE ON PTT SCORES

	Source		
	df	MS	F
Experiment I			
Instructions (A)	1	.0	
Heuristics (B)	1	.16	4.00*
A $\times$ B	1	.0	—
Error	84	.04	
Experiment II			
Instructions (A)	1	.06	1.00
Word Association (B)	2	.34	5.67**
A $\times$ B	2	.015	—
Error	80		

Note.—This analysis was done by the method for unequal *N*s presented by Walker and Lev, 1953, pp. 381-382.

* $p < .05$.

** $p < .01$.

measures, the question posed above cannot be answered satisfactorily from these data. The possible contribution of fluency to performance will remain an open question, particularly on a test such as the UUT which does not permit fluency measurement, unless future designs provide a suitable measure and control for fluency.

DISCUSSION

The chief hypothesis stated that task-relevant principles exist for creative thinking tasks, and that when made available to Ss these principles will have a facilitating effect on performance. That strategies are relevant to performance on the UUT is indicated by the report of controls that they used strategies for many of their uses. However, the finding that heuristics-trained groups used strategies successfully significantly more than controls suggests that strategies are employed in a more intuitive, random fashion prior to training, and that after training strategies are used more deliberately and efficiently.

The findings on the second hypothesized effects confirm the expectation that word-association training can enhance performance, but they yield no evidence of greater effects as a function of length of training. Changes made in the design were such that we believed that a better test of the length of training variable would result; consequently we have little confidence in this variable. Maltzman's theory that word-association training produced mediated generalization from uncommon responses in one hierarchy to those in another is further attenuated by the observation that Ss often employed strategies, since it would be difficult to explain the simultaneous occurrence of strategies and mediated generalization. Chiefly, however, the evidence supplied by

Gallup (1963) is damaging to an S-R interpretation. If there are grounds for Gallup's suspicion that word association produced a set to respond *divergently*, then different responses could be predicted without using S-R principles at all.

That heuristics training involved more than strategies is evident from the facilitation of the PTT, a task for which strategies are not relevant. It may be that training of any kind will at least partially involve producing a differential in set. Word association is not necessarily an especially efficient method of achieving such a set.

The findings with respect to the other hypothesized effects add to our picture somewhat. Instructions to be original evidently were perceived as specific to the UUT. Although instructions quite possibly contribute to a set to respond divergently, it is more important to notice that a person can adjust his performance to be more original when it is clear that originality is called for. Significant differences in the degree of association between measures of originality and fluency, for experimental versus control Ss, suggests that a set produced by either type of training makes Ss use their resources more efficiently.

In summary, we would say that Ss acquired a set to respond according to the instructions and requirements of the task, and often used relevant principles to so respond. These principles can be made more accessible to conscious, systematic use if they are unambiguously stated and illustrated for the S. This was accomplished by heuristics treatments, while word-association treatments primarily influenced set and left the use of principles random and inexplicit. Influence on set, however, may make Ss more efficient in fulfilling the task. These

effects of set are not specific to a task, but they are probably of short-term effectiveness.

The more important question to be faced is whether we have facilitated originality at all. Although it is absurd to claim that we have influenced creativity in 20 minutes of training, our results may point to an operational method of actually training originality. This method of training would confront Ss with principles for several tasks, by way of inducing Ss to search for strategies on new tasks *without* training. The first step would be to formulate strategies for several tasks. If strategies are relevant to the post-test in this training, we would expect Ss to be more sensitive to their relevance than controls.

This line of thought suggests that we try conceiving of the creative individual as one who is able, when faced with a particular creative thinking task, to find the relevant principles and apply them. The primary appeal of this conception seems to be that the criterion of creativity is linked firmly with performance, yet preserves an important tie with the actual process of creation. The hypothesis of deeper levels of the creative process, as stated by Gordon (1961) is altogether plausible, but perhaps better set aside at this point until operational

training has been explored as far as possible.

REFERENCES

- CHRISTENSEN, P. R., GUILFORD, J. P., & WILSON, R. C. Relations of creative responses to working time and instructions. *Journal of Experimental Psychology*, 1957, **53**, 82-88.
- GALLUP, H. Originality in free and controlled association responses. *Psychological Reports*, 1963, **13**, 923-929.
- GORDON, W. J. J. *Synectics, the development of creative capacity*. New York: Harper, 1961.
- GUILFORD, J. P. Creativity. *American Psychologist*, 1950, **5**, 444-454.
- GUILFORD, J. P. Creative abilities in the arts. *Psychological Review*, 1957, **64**, 110-118.
- MALTZMAN, I. On the training of originality. *Psychological Review*, 1960, **67**, 229-242.
- MALTZMAN, I., SIMON, S., RASKIN, D., & LICHT, L. Experimental studies in the training of originality. *Psychological Monographs*, 1960, **74**(6, Whole No. 493).
- OSBORN, A. F. *Applied imagination*. (2nd ed.) New York: Scribners's, 1957.
- ROSENBAUM, M. E., ARENSON, S. J., & PANMAN, R. A. Training and instructions in the facilitation of originality. *Journal of Verbal Learning & Verbal Behavior*, 1964, **3**, 50-56.
- SIMON, S. Effects of training on word association originality and unusual uses. Unpublished doctoral dissertation, University of California, 1961.
- WALKER, H., & LEV, J. *Statistical inference*. New York: Henry Holt, 1953.
- WILSON, R. C., GUILFORD, J. P., & CHRISTENSEN, P. R. The measurement of individual differences in originality. *Psychological Bulletin*, 1953, **50**, 362-370.

(Received June 13, 1966)

EFFECT OF INTRINSIC AND EXTRINSIC REINFORCEMENT CONTINGENCIES ON LEARNER PERFORMANCE¹

HOWARD J. SULLIVAN, ROBERT L. BAKER, AND RICHARD E. SCHUTZ
Arizona State University

76 Air Force ROTC cadets participated in an investigation of the effects on learner performance of variations in (a) knowledge of results for responses to mastery items inserted in instructional material and (b) reinforcement contingencies for performance on a 100-item final criterion test over the material. Ss receiving no knowledge of results for mastery-item responses scored significantly higher ($p < .001$) on the 11 mastery tests than did Ss receiving immediate feedback through chemically treated answer sheets. However, the 2 groups did not differ significantly in criterion-test performance. $\frac{1}{2}$ of the Ss could earn \$4 for a criterion-test score of 80% or higher, while the remaining Ss earned \$2.50 irrespective of criterion-test performance. No significant differences in test scores were associated with these reinforcement conditions. The data suggest that Ss receiving immediate feedback employ a markedly different strategy from "no feedback" Ss in learning instructional material.

The facilitating effect of reinforcement on student learning is well recognized. Yet in studies of applied learning the term "reinforcement" frequently has been used to describe a variety of stimulus conditions without specifying the possible differential effects related to these diverse conditions. For example, in many studies knowledge of results for student responses to en route test items over the instructional materials has been treated as reinforcement. Reinforcement in these studies is a condition that is intrinsic to (i.e., built into) the learning materials. In other studies reinforcement is a stimulus condition extrinsic to the learning materials. Here, the presentation of reinforcement is contingent upon level of performance on the instructional material, but the reinforcement is not a condition built directly into the material. Gold stars, course grades, and special awards for

performance serve as examples of extrinsic reinforcement.

The present study sought to investigate the effects of both intrinsic and extrinsic reinforcement on student performance when both conditions were employed in the same instructional program. The intrinsic reinforcement condition included variations in the knowledge of results provided for student responses to sets of mastery items inserted at various points in the instructional material. The extrinsic reinforcement condition consisted of variations in the amount of money that could be earned by subjects (Ss) for acceptable performance on a final criterion test over the material.

METHOD

Subjects

The Ss were 76 Air Force Reserve Officers Training Corps cadets who were second semester freshmen enrolled in the AFROTC program at Arizona State University. The Ss were selected at random from among approximately 100 cadets who volunteered to participate in the experiment during the time when they normally attended AFROTC classes and drill periods.

¹ This research was supported under Contract AF 33 (615) 1507 with the Aerospace Medical Research Laboratories, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio.

Design

The study employed a 2×2 factorial design. The variations in intrinsic reinforcement conditions served as one factor in the design. The variations in extrinsic conditions served as the second factor.

Procedures

The Ss attended four 50-minute class periods on a twice-a-week basis to read the instructional material and answer the mastery items. Each S was randomly assigned either to a room in which knowledge of results was provided for responses to mastery items or to a room in which no knowledge of results was provided. The extrinsic reinforcement conditions were randomized within both the "feedback" and "no feedback" rooms. Printed instructions explaining the appropriate extrinsic reinforcement contingency and specifying the time schedule for the experiment were given to each S at the beginning of the first class meeting. The final criterion test was administered to all Ss at a fifth session 2 days after the final instructional period.

The Ss read the instructional material and responded to the mastery tests at their own pace. The "feedback" group received immediate feedback on their responses to each mastery item as a function of chemically treated answer blanks. Individuals in this group used special pens to mark their responses to the mastery items. When S marked the correct response blank, the blank turned red; when he marked an incorrect blank, it turned yellow. The Ss in this group were told that if their first response to an item was incorrect, they were to continue responding to the item until they answered it correctly. The "no feedback" group received no knowledge of results on their responses to the mastery items.

Two levels of extrinsic reinforcement were included in the study. Instructions for Ss under the Contingent Reinforcement condition stated that S would be paid \$4.00 if he scored 80% or higher on the final test over the instructional material, \$2.00 if he scored from 50% to 79%, and nothing if he scored below 50%. Individuals in the Assured Reinforcement group were told that they would be paid \$2.50 if they attended all five scheduled sessions. Instructions to the Assured Reinforcement group stated that there would be a final test over the instructional material, but the instructions did not relate the test to the \$2.50 in any way. Thus, no extrinsic reinforcement based upon level of

performance on the criterion test was available to the Assured Reinforcement group.

Materials

The textual material used in the experiment was a revised edition of the Air Force manual, *The Military Justice System* (United States Air Force Reserve Officers Training Corps, 1962). This text was originally selected for revision because it included sufficient concept complexity and developmental continuity directly applicable to Air Force and ROTC curriculum. After subjecting the original version of the text to a logical analysis of objectives, a final list of 69 specific behavioral objectives was compiled. The Instructional Specification strategy (Schutz, Baker, & Gerlach, 1964; Schutz, Baker, & Sullivan, 1966) was then employed to specify the stimulus conditions required for the attainment of these objectives. The materials were revised on the basis of these specified conditions, item analyses of the performance of cadets from earlier studies on the final criterion test, analyses of interview data from individual cadets who had read the materials, and application of the gap and mastery principles of programmed instruction (Silberman, Coulson, Melaragno, & Newmark, 1964).

A set of 131 mastery items was developed pertaining to the behavioral objectives specified for the instructional materials. These items were grouped into 11 unit-mastery tests and inserted at appropriate points in the 60-page revised text. Each mastery test contained items covering only the material from the section of the text immediately preceding the test. Determination of the place in the text where each mastery test was inserted was a function of optimum length or logical determination of appropriate homogeneous blocks of content.

A final criterion test of 100 three- and four-choice multiple-choice items was developed from an original pool of 200 items. Only items with a difficulty index of .75 or lower for a sample of 50 Ss who had read the original text were included in the final 100-item test. The reliability coefficient for the criterion test, computed by the KR-20 formula on a sample of 76 Ss who had read the revised text, was .86.

RESULTS

The mean score for each group on the 100-item criterion test is shown in Table 1. It is apparent from the table

that the Contingent Reinforcement group scored approximately three points higher than the Assured Reinforcement group, and the "no feedback" group scored slightly more than one point higher than the "feedback" group. These differences were tested for significance using a two-way analysis of variance. Neither the main effects nor the interaction was statistically significant.

Analyses of performance on the mastery tests, however, did reveal important differences between the treatment groups. Only three of the 38 Ss in the "feedback" group failed to complete all 11 mastery tests during the four periods of the instructional program. In the "no feedback" treatment, however, 15 of the 38 individuals completed only 10 or fewer tests and failed to reach the final mastery test. Clearly, the "no feedback" group was spending more time studying either the textual material or the mastery items than was the "feedback" group.

Descriptive statistics relating to the mean standard score on unit-mastery tests completed by each S are shown in Table 2. Since data on mastery-test scores were not available on four Ss from the contingent-reinforcement-plus-feedback cell when the statistical analyses were performed, these Ss were assigned the computed mean

TABLE 2
MASTERY TEST MEAN STANDARD SCORES

Extrinsic reinforcement	Intrinsic reinforcement				Totals	
	Feedback		No feedback			
	N	Stand-ard score	N	Stand-ard score	N	Score
Contingent	19	47.10	19	53.24	38	50.17
Assured	19	47.07	19	51.97	38	49.52
Totals	38	47.09	38	52.61	76	49.85

standard score for their cell on mastery-test performance. This accounts for the slight variation of the grand mean score (49.85) from a grand mean of 50.00.

The data in Table 2 reveal that the "no feedback" Ss performed considerably better on the mastery items than did the "feedback" Ss. The mean standard score for mastery-test performance is 5.52 standard score points higher for the "no feedback" group. A two-way analysis of variance of mastery-test scores revealed that this difference is significant at the .001 level ($F = 22.08$). Neither the extrinsic reinforcement contingency effect nor the interaction between intrinsic and extrinsic reinforcement approached statistical significance.

DISCUSSION

The results of the study suggest that there were important differences between the "feedback" and "no feedback" treatment groups in the strategies that they employed to learn the instructional material. The better performance of the "no feedback" group on the mastery tests and the failure of 15 Ss from this group to finish the instructional program indicate that the "no feedback" Ss expended more time and effort attempting to learn from the prose textual material. For

TABLE 1
CRITERION TEST MEAN SCORES

Extrinsic reinforcement	Intrinsic reinforcement				Totals	
	Feedback		No feedback			
	N	Score	N	Score	N	Score
Contingent (\$4-\$2.0)	19	61.37	19	61.32	38	61.35
Assured (\$2.50)	19	57.00	19	59.63	38	58.32
Totals	38	59.18	38	60.47	76	59.83

these individuals, of course, this is the only instructional material in the text. The "feedback" Ss, on the other hand, apparently neglected the textual material to some degree and used the instructional value of the immediate feedback to their mastery-item responses. Such a procedure would account for their greater speed in working through the textual material and their inferior performance on the mastery items. That the "feedback" Ss were successful in learning from the immediate feedback to their mastery-test responses is demonstrated by their subsequent performance on the criterion test. Their criterion-test performance was comparable to that of the "no feedback" group, even though their mastery-test performance was significantly inferior.

An interesting phenomenon to note here is the apparent sensitivity of the learner to subtle procedural cues implicit in the instructional material. For example, one might predict that both the "feedback" and "no feedback" groups would study equally hard on the textual material and that no difference would occur in mastery-test performance between the two groups. Since feedback is not received until after the learner responds to a mastery item, one would expect that on subsequent items over the same material (e.g., the criterion test) the feedback would result in an advantage for the individuals receiving it. However, it appears that "feedback" Ss quickly observe that they need not labor over the textual materials to learn the material to be covered on the criterion test. Where individuals in the "no feedback" group may choose to look on the preceding pages for the correct answer to a puzzling mastery item, an S in the "feedback" group can employ the easier and simpler expedient of marking in succession his

highest-order response choice until the feedback indicates a correct response.

How can one capitalize upon the advantages of the intrinsic reinforcement involved in the immediate feedback procedure on the mastery items while at the same time maintaining the control of the textual material over the reader's learning? One possible procedure would be to provide the learner with extrinsic reinforcement for acceptable performance on the mastery items, as well as for acceptable performance on the criterion test. Thus, performance of the "feedback" Ss on the mastery test should be improved because of the extrinsic reinforcement associated with good mastery performance. The immediate feedback on these items should still serve to facilitate subsequent performance on the criterion test.

A final word should be said concerning the effect of extrinsic reinforcement in using instructional materials. The differences in the levels of the monetary reinforcement contingency employed in the present study were not of sufficient strength to significantly affect student performance. However, there is little doubt that extrinsic reinforcement is required in the learning task to maintain control of the instructional material over student responses. In the classroom setting such teacher strategies as the use of appropriate verbal statements (praise, encouragement, exhortation, etc.), assignment of perseverant students to preferred activities, and the permission of free choice of student activity upon completion of assigned work may be employed to develop and maintain desired learner responses to learning materials. The use of these strategies and other effective procedures by the classroom teacher is an essential technique for maximizing the effectiveness of instructional materials.

No matter how excellent the quality of the material, the student will not learn it well unless he is provided with an incentive for doing so.

REFERENCES

SCHUTZ, R., BAKER, R., & GERLACH, V. Measurement procedures in programmed instruction. Project No. 909, Title VII National Defense Education Act Grant 7-12-0300-160, 1964.

SCHUTZ, R., BAKER, R., & SULLIVAN, H. Instructional components for behavioral objectives. Paper read at American Educa-

tional Research Association, Chicago, February 1966.

SILBERMAN, H., COULSON, J., MELARAGNO, R., & NEWMARK, G. Use of exploratory research and individual tutoring techniques for the development of programming methods and theory. Title VII National Defense Education Act Grant 7-14-000-181, 1964.

UNITED STATES AIR FORCE RESERVE OFFICERS TRAINING CORPS. *The military justice system*. Montgomery, Alabama: Air University, 1962.

(Received June 20, 1966)

PREDICTION OF ADOLESCENT POPULARITY AND REJECTION FROM ACHIEVEMENT AND INTEREST TESTS¹

HERBERT HOROWITZ²

University of Pittsburgh

15 Project TALENT scores for 1,437 male and 1,505 female high school students were correlated with popularity and rejection scores achieved with members of their own and the opposite sex. It was found that: (a) multiple correlations enhance considerably the predictability of such sociometric choices; (b) generally, popularity criteria were more correlated with predictors than rejection criteria; (c) best predictors of both popularity and rejection were English test total, information about and interest in sports, and socioeconomic status; and (d) some variables seem to relate only to popularity or only to rejection. The implications of the latter finding for use of multivariate analytic methods in isolating clusters of variables associated with popularity and/or rejection were discussed briefly.

In the 35 years since its introduction by Moreno, the technique of sociometric measurement has come to be used widely for studying interpersonal choice. A major trend in this research has been the effort to isolate factors associated with popularity and rejection among children and adolescents. Summaries of the research findings in this area, as well as in other areas of sociometric measurement, are contained in the reviews of Lindzey and Borgatta (1954) and in *The Sociometry Reader* (Moreno et al., 1960). To cite but a few of the large body of related findings, it has been found, for example, that popularity is positively associated with athletic ability and interest (Feinberg; 1953; Feinberg, Smith, & Schmidt, 1958; Krumboltz, Christal, & Ward, 1959;

McGraw & Tolbert, 1953), positively associated with sociability and social adjustment (Bonney, 1946; Feinberg, 1953; Feinberg et al., 1958; Kuhlen & Bretsch, 1960), and negatively associated with school newspaper activity (Krumboltz et al., 1959).

The present study sought to extend these findings using multiple rather than single criteria typical of previous work in the area and using a larger and more representative sample of adolescents than has been used previously.

METHOD

Description of the Sample

The sociometric data for the present study were collected in conjunction with the high school testing program constituting Project TALENT (Flanagan, Davis, Dailey, Shaycoft, Gorham, Orr, & Goldberg, 1964). The sociometric data came from a subsample of eight schools selected from the 1,353 schools constituting the TALENT sample. The eight schools were chosen to maximize generalizability to the national high school age group. Each school in the sample was representative of one of the eight (continental) regional areas defined by the United States Office of Education. In selecting schools for the subsample, however, an attempt was made to limit the variability of

¹This research was supported by the United States Office of Education, Cooperative Research Branch, under Contract OE 3-10-116 with the University of Pittsburgh.

²This research was done while the author was also affiliated with the American Institutes for Research, Washington, D.C. The author is now at Wake Forest College. The author is indebted to Edwin A. Locke for many helpful comments on earlier versions of this paper.

the size of the schools. Thus, the number of students in each grade ranged only from 57 to 161. Summing across all grades and all schools, the sample contained 1,437 males and 1,505 females. For all of these individuals sociometric data as well as regular Project TALENT testing data were available.

The Sociometric Criterion Scores

Within each of the four grades, each male and female student was asked to list the three boys within their own grade they most "like to be with" and, in addition, the three boys they least "like to be with." Each student was then required to give the same six nominations (three positive and three negative) for the female students in his or her class. Thus, four kinds of sociometric scores were obtained for each student.

1. Same-Sex Attraction Score (SS-A): the number of times the individual was nominated as liked by members of his (her) own sex, weighted by dividing by the number of same-sex students in that grade.

2. Opposite-Sex Attraction Score (OS-A): the number of times the individual was nominated as liked by members of the opposite sex, weighted by dividing by the number of opposite-sex students in that grade.

3. Same-Sex Rejection Score (SS-R): the number of times the individual was nominated as disliked by members of the same sex, weighted by dividing by the number of same-sex students in that grade.

4. Opposite-Sex Rejection Score (OS-R): the number of times the individual was nominated as disliked by members of the opposite sex, weighted by dividing by the number of opposite-sex students in that grade.

The actual weights used in computing these scores varied from the low 20's to the upper 80's. From this it can be seen that the social groups in which these scores were taken, while variable, were not very large.

Product-moment correlations among these criteria in each sex showed the two kinds of attraction scores (SS-A and OS-A) to be highly positively correlated, as were the two kinds of rejection scores (SS-R and OS-R). Furthermore, correlations between attraction and rejection scores, while significantly different from zero, were much lower and negative. Factor analyses³ of these criteria

in each sex (reported in Horowitz, 1966) yielded two independent factors, readily identified as "attraction" (same-sex and opposite-sex) and "rejection" (same-sex and opposite-sex). These factors accounted for approximately 80% of the total variance in the criterion scores. Nevertheless, it was decided to use the four criterion scores separately in the present analyses since it was possible to obtain variations between the two kinds of attraction criteria or between the rejection criteria in their relationships to the predictors.

The Predictors

In all, 15 Project TALENT variables were used as predictors.⁴ These variables (listed in Table 1) were the ones remaining after preliminary analyses with a larger group of predictor variables resulted in elimination of variables whose relations to the criteria had been shown to be weak or inconsistent. Although such variable-selection procedures have been employed previously (e.g., Krumboltz, Christal, & Ward, 1959), there is always the danger of inadvertently capitalizing on chance relationships in the data. However, this does not appear to be the case here. In the original matrices of correlation of 56 predictors and 4 criteria (sexes combined in these matrices), 158 (70%) of the 224 correlations were significant at the $p = .05$ level, as compared with 11 correlations which would be significant if only chance conditions were operating.⁵ Thus, it is clear that many variables other than the 15 selected were related to the criteria. Although the original analyses were done with combined sexes, the analyses to be reported here were done separately for each sex. In this way, it was possible to study the replicability of the findings across sex.

In the analysis each of the 15 predictors was correlated individually with the four criteria. In addition, multiple correlations were performed with composite sets of the 15 predictors.

RESULTS

The upper portion of Table 1 shows the intercorrelations among 15 predic-

⁴Standardization and other normative data about all 15 predictors (except the socioeconomic index) are contained in Flanagan, Davis, Dailey, Shaycoft, Orr, Goldberg, and Newman, 1964.

⁵With samples of this size, r s as small as $\pm .06$ are significantly different from zero at the $p = .05$ level.

³Principal components solution with units in the diagonal followed by normalized varimax rotation.

TABLE 1
PRODUCT-MOMENT AND MULTIPLE CORRELATIONS AMONG FIFTEEN PROJECT
TALENT PREDICTOR VARIABLES AND FOUR SOCIOMETRIC CRITERIA
BY SEX

Predictors	Intercorrelations							
	Males				Females			
	Attractiveness		Rejection		Attractiveness		Rejection	
	SS-A	OS-A	SS-R	OS-R	SS-A	OS-A	SS-R	OS-R
Literature information	.12	.14	-.01	-.06	.06	.08	-.09	-.07
Music information	.08	.19	-.01	-.08	.11	.12	-.12	-.12
Scientific attitude information	.12	.14	-.02	-.06	.11	.16	-.14	-.14
Sports information	.22	.27	-.11	-.15	.09	.13	-.09	-.10
Sociability scale	.17	.22	-.04	-.06	.15	.18	.03	-.03
Impulsiveness scale	.01	.00	.09	.07	-.00	.04	.10	.06
Leadership scale	.16	.20	.02	-.01	.15	.22	.03	-.01
Literary-linguist interest	.06	.14	.06	.02	.04	.11	-.02	-.03
Social service interest	.04	.12	-.01	-.06	.10	.10	-.08	-.08
Musical interest	.02	.11	.09	.03	.04	.04	-.01	.01
Sports interest	.18	.17	-.15	-.19	.12	.15	-.04	-.09
Outdoors interest	.01	-.03	-.06	-.12	.03	.07	-.00	-.02
Labor interest	-.08	-.13	.01	.01	-.05	-.07	.02	.04
English test total	.16	.21	-.09	-.12	.18	.16	-.18	-.18
Socioeconomic level index	.11	.17	.03	-.04	.12	.19	-.08	-.12
Corrected R_{mult}	.31*	.39*	.25*	.27*	.29*	.34*	.26*	.26*

Note.—For samples of this size to be significant at the $p = .01$ level, r must be greater than or equal to .08. For males, $N = 1,437$; for females, $N = 1,505$.

* $p < .01$.

tors and 4 criteria. It can be observed that, in absolute terms, the correlations are small, none of them exceeding $\pm .30$, and many do not exceed $\pm .10$. Second, the criteria of attractiveness (SS-A and OS-A) are generally more strongly correlated with the predictors than the corresponding criteria of rejection. This is particularly true for the males. A third point to note is that the correlations show a reflection in sign from those in the attractiveness columns (usually positive) to those in the rejection columns (usually negative).

The bottom row of Table 1 shows the eight multiple correlations between each of the criteria and the 15 predictors. All of these correlations are statistically significant and account

for at least twice as much variance as the largest bivariate correlations between that criterion and any of the predictors. That is, there is considerable enhancement of predictability of both popularity and rejection criteria through multiple correlation technique.

The following variables were the best predictors of popularity in both sexes: the English test total, interest in sports, and the self-rating personality scales of sociability and leadership. A particularly important additional variable for predicting males' popularity is their *knowledge* of sports (sports information test).

The results for rejection are, in some respects, the reverse of those for popularity. For both males and fe-

males, English test total and information about and interest in sports are negatively related to rejection; and, for females, the socioeconomic variable predicts both popularity (positive association) and rejection (negative association). From these results it is clear that some overlap exists between factors associated with popularity and those accounting for rejection. To see *how much* overlap exists, the obtained correlations for attractiveness were correlated with each other and with those for rejection. In doing this, the algebraic sums of the two product-moment correlations for each predictor in each adjacent pair of columns were obtained (SS-A + OS-A and SS-R + OS-R). Then these four sets of 15 sums of correlations were themselves correlated. These *r*s are shown in Table 2. The table shows that, within a given kind of affect, cross-sex correlations of the sums of original correlations are positive and high (particularly for attractiveness). In this way, these results provide some proof of replicability of the findings for independent subsamples. However,

more importantly, the correlations between attractiveness and rejection (both within and across sexes) are also high and, predictably, negative. Thus, these data show clearly that the *patterns of association* among predictors and criteria are quite consistent, that is, that the *relative* power of different predictors remains quite stable with variations in sex of *S* and nature of affect.

DISCUSSION

The present data delineate a pattern of interpersonal values in the adolescent world of American high schools which is similar to the one described by Coleman in his extensive study of *The Adolescent Society* (Coleman, 1961). One of the major findings of that report was that, although athletes were chosen more frequently (by several criteria of interpersonal popularity) than scholars, the "athlete-scholars" were the most popular of all, suggesting positive valuation of both academic and athletic achievements. The present results support that finding in showing strong relationships between, on the one hand, popularity and rejection criteria and, on the other hand, interest and achievement predictors from both the intellectual and athletic domains.

While the present study supports previous investigations in showing a multiplicity of associations between sociometric choices and other factors, clearly the variables were not equally successful as predictors of popularity and rejection. While some overlap exists among variables associated with popularity and those associated with rejection, these variables (English test total, sports information, sports interest, socioeconomic index) are a minority of the complete set of 15. The finding that relatively few variables relate to both popularity and

TABLE 2

INTERCORRELATIONS OF FOUR SETS OF
FIFTEEN CORRELATION COEFFICIENTS
BETWEEN FIFTEEN PROJECT TALENT
PREDICTORS AND SOCIOMETRIC
CRITERIA

	Attractiveness		Rejection	
	Males	Females	Males	Females
Attractiveness				
Males	—	+.87	-.70	-.62
Females	—	—	-.53	-.56
Rejection				
Males	—	—	—	+.64

Note.—All correlations in this table are significantly different from zero at or beyond the $p = .05$ level. See text for explanation of derivation of correlations reported in this table.

rejection is meaningful in light of the results obtained with the criterion score analyses (Horowitz, 1966). The fact that the criteria separated into two factors, "attraction" and "rejection," suggests that predictors related to attractiveness might be different from those related to rejection. Support for this expectation was found in the present data, in spite of the significantly consistent patterns of prediction which were also found (Table 2). For instance, leadership and sociability personality scales have fairly strong positive associations with popularity, but, with one exception, statistically nonsignificant relationships to rejection. Similarly, while labor interest was negatively related to attractiveness, it had only nonsignificant correlations with rejection. Finally, at least one of the 15 variables, the impulsiveness scale, seemed to be associated only with rejection.

Thus, the present data, when considered alongside the results of the criterion analysis, suggest that there may be three clusters of psychometric variables which predict popularity and rejection: those associated only with popularity, those associated only with rejection, and those associated (to a lesser degree) with both popularity and rejection. The ultimate test of this hypothesis would require a multivariate analytic procedure with a different, and probably much larger, set of predictors which had previously been shown to be related to popularity and/or rejection. Such a procedure would not only increase the total variance which multiple correlations could explain; it would also reveal the particular constellations of variables which are associated, in the first place, with popularity; second, with rejection;

and, finally, with both popularity and rejection.

REFERENCES

- BONNEY, M. E. A sociometric study of the relationship of some factors to mutual friendships on the elementary, secondary, and college levels. *Sociometry*, 1946, **9**, 21-47.
- COLEMAN, J. S. *The adolescent society*. Glencoe, Illinois: The Free Press, 1961.
- FEINBERG, M. R. Relation of background experience to social acceptance. *Journal of Abnormal and Social Psychology*, 1953, **48**, 206-214.
- FEINBERG, M. R., SMITH, M., & SCHMIDT, R. An analysis of expressions used by adolescents at varying economic levels to describe accepted and rejected peers. *Journal of Genetic Psychology*, 1958, **93**, 133-148.
- FLANAGAN, J. C., DAVIS, F. B., DAILEY, J. T., SHAYCOFT, M. F., ORR, D. B., GOLDBERG, I., & NEWMAN, C. A., JR. The American high school student. United States Office of Education Cooperative Research Branch Technical Report No. 635. Pittsburgh: University of Pittsburgh, 1964.
- HOROWITZ, H. Interpersonal choice in American adolescents. *Psychological Reports*, 1966, **19**, 371-374.
- KRUMBOLTZ, J. D., CHRISTAL, R. E., & WARD, J. H., JR. Predicting leadership ratings from high school activities. *Journal of Educational Psychology*, 1959, **50**, 105-110.
- KUHLEN, R. G., & BRETSCH, H. S. Sociometric status and personal problems of adolescents. In J. L. Moreno et al. (Eds.), *The sociometry reader*. Glencoe, Illinois: The Free Press, 1960.
- LINDZEY, G., & BORGATTA, E. F. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology*, Vol. I. Reading, Massachusetts: Addison-Wesley, 1954.
- MCCRAW, L. W., & TOLBERT, J. W. Sociometric status and athletic ability of junior high school boys. *Research Quarterly of the American Association of Health, Physical Education, and Recreation*, 1953, **24**, 72-80.
- MORENO, J. L., et al. (Eds.), *The sociometry reader*. Glencoe, Illinois: The Free Press, 1960.

(Received June 29, 1966)

TEACHERS' ATTITUDES TOWARD CHILDREN'S BEHAVIOR REVISITED

ALEXANDER TOLOR, WILLIAM L. SCARPETTI

Fairfield University

AND PAUL A. LANE

University of Bridgeport

This study concerned itself with the evaluations of a wide range of child behavior by 118 elementary school teachers and 23 clinical psychologists. The Staten Island Behavior Scale was administered to these two groups with the instruction that each of the 295 behavioral items be rated as indicating either normal or abnormal behavior. The items were also grouped into 8 classifications on the basis of a consensus by a panel of judges. Teachers were found to differ from psychologists in their attitudes toward child behavior on a significant 22% of the items, mostly falling in the categories aggressive behavior, regressive behavior, and affect expression. The teachers tended to regard almost all of the differentiating behavioral descriptions abnormal whereas the psychologists perceived them to be normal. Teacher experience was found to be a significant variable with inexperienced teachers differing more from the psychologists, and ascribing more pathology to a variety of child behavior, than experienced teachers. The findings were related to the "continuity hypothesis" and to the difference in roles between teachers and clinicians.

A classic study by E. K. Wickman (1928) found a great discrepancy between the views of teachers and mental health workers toward the behavior problems of children. Although this study has been criticized on methodological grounds, its influence on American education has been profound. Several more recent investigations (Griffiths, 1952; Schrupp & Gjerde, 1953; Stouffer, 1952) suggest that while there has been considerable change in the attitudes of teachers to make them more congruent with those of clinicians, marked differences between the two groups continue to exist. These differences are still in the direction of teachers being more concerned with management, sexual adjustment, and adherence to authority problems whereas the mental health professionals are more sensitive to withdrawal behavior and behavior not directly related to the school routine but suggesting a deterioration of social or emotional patterns.

The purpose of this study was to explore further the relationship between the evaluations by teachers and psychologists of a wide range of child behavior patterns. More specifically, it was anticipated that by employing a comprehensive scale of unambiguous behavioral items, which can be grouped on an empirical or theoretical basis to focus on patterns of functioning, it would be possible to identify the types of behavioral patterns that teachers and psychologists perceive most differently. The effect that the teachers' experience level has on the ratings was also to be determined.

METHOD

The teacher respondents consisted of 90 female and 28 male elementary public school teachers randomly selected from a large urban school system. They were drawn from all grades ranging from the kindergarten to the seventh grade level, inclusively. There were 9 teachers at the kindergarten level, 13 at Grade 1, 16 at Grade 2, 15 at Grade 3, 17 at Grade 4, 16 at Grade 5, 15 at Grade 6, and 17 at Grade 7. The profes-

sional experience of the teachers encompassed the range of less than 1 year to 44 years (mean 11.9 years).

The psychologist respondents consisted of 17 males and 6 females, all of whom were functioning in clinical settings in the same state as the teachers. The highest degree held by the psychologists was the Ph.D. for 15 and the M.A. or M.S. by eight. In experience the psychologists ranged from under one year to over 30 years (mean 10.7 years).

The measuring device was the Staten Island Behavior Scale (Mandell & Silberstein, 1965) which consists of 295 items descriptive of children's behavior. The items were originally selected from published and unpublished scales used to evaluate children's adjustment and from an analysis of a large number of case records in the files of a child guidance clinic.

The items were classified for the purposes of the present study by six raters (5 advanced students and 1 Diplomate in Clinical Psychology), making independent judgments, into the following classifications: psychosomatic and physical disturbance (71 items); phobias (18 items); aggressiveness (56 items); affect expression (58 items); communication disturbance (21 items); regressive behavior (15 items); inefficiency indicators (25 items); and fantasy involvement or withdrawal (31 items).

Each item was placed into the category that represented the rating of the majority of the judges. An indication of the high degree of interrater agreement is provided by the fact that on 203 of the 295 items (69%) at least five of the raters were in complete agreement in regard to the classification of an item.

Illustrations of the different types of items are the following: For the psychosomatic and physical disturbance classification—"Is slow in his movements"; for the phobic classification—"Is afraid of being alone in a wide open space"; for aggressiveness—"Hits or attacks other child"; for affect expression—"Shows inappropriate feeling"; for communication disturbance—"Talks and talks"; for regressive behavior—"Carries blanket"; for inefficiency indicators—"Does not complete his chores"; and for fantasy involvement or withdrawal—"Doesn't join in competitive games."

The written instructions accompanying the administration of the scale are presented below:

For each of the following items please indicate whether the behavior in question,

in your opinion, indicates normal or abnormal behavior in a child falling in the age range from 1 to 16 inclusive.

Please answer all items without omitting any, and try to check either the "Normal" or "Abnormal" category. In the event that you really cannot decide whether the behavior is normal or abnormal, you may check the "Unknown" line. However, you will probably be able to arrive at a definite decision in all or nearly all instances.

The respondents were not given any time limit, but were cautioned against collaborating with anyone else in completing the scale.

The very broad age range was quite deliberately employed in the instructions since our intent was not so much to obtain reactions to a child's behavior at a specific point in time—even though we recognized that the appropriateness of behavior is age related—but more to distill from a less structured frame of reference behavior patterns that most frequently tend to be regarded as being normal or abnormal, even in the absence of a specific anchoring point.

RESULTS AND DISCUSSION

The main findings indicated that teachers and psychologists, when their responses for each item on the questionnaire are compared by chi-square, differ significantly ($p < .05$) on 66 of the 295 items, that is, on 22.4% of the items, in regard to whether they rated the behavior to be normal or abnormal. This number of differentiating items is significantly greater than would be expected on a chance basis alone. Of the 66 critical descriptions of behavior, 7, or 11%, are in the psychosomatic and physical disturbance category, 1, or 2%, in the phobias category, 21, or 32%, in the aggressiveness category, 19, or 29%, in the affect expression category, 3, or 5%, in the communication disturbance category, 6, or 9%, in the regressive behavior category, 3, or 5%, in the inefficiency indicators category, and 6, or 9%, in the fantasy involvement or withdrawal category.

Since the categories consisted originally of unequal numbers of scale items, the percentages reported above may be somewhat misleading. Another way of analyzing the same data is to determine the proportion of items within each category that differentiates the teachers' judgments from the psychologists' ratings. Employing this approach, we note from Table 1 that the greatest disagreement occurs in the areas of regressive behavior, aggressiveness, and affect expression. Next in order are fantasy involvement or withdrawal, communication disturbance, and inefficiency indicators in which areas the two groups are in relatively close agreement. In regard to phobias and psychosomatic and physical disturbance, the judgments of psychologists and teachers are very much in accord, as can be seen by the negligible item disagreement.

These results indicate, therefore, that elementary school teachers in general tend to evaluate behavior that may be described as regressive, aggressive, and emotional quite differently than do psychologists. In view of the fact that nearly all of the differentiating items, that is, 61 of 66, or 92.4%, were rated predominantly normal by psychologists and abnormal by teachers, it is obvious that elementary school teachers perceive regressive, aggressive, and emotional behavior to be considerably more pathological than do mental health professionals.

Two subgroups of teachers were selected based on amounts of teaching experience. The Highs consisted of the top third in experience of the overall group of teachers. The Lows consisted of the lowest third in teaching experience. The 39 teachers in the High subgroup ranged in professional experience from 14 to 44 years with a median of 24.5 years; the 39 teachers

TABLE 1
DEGREE OF DISAGREEMENT BETWEEN
TEACHERS AND PSYCHOLOGISTS IN
SPECIFIC SCALE CATEGORIES

Category	Items in scale	Percentage of items disagreed upon
Physical-psychosomatic	71	10
Phobic	18	6
Aggressive	56	38
Affect	58	33
Communication	21	14
Regressive	15	40
Inefficiency	25	12
Fantasy-withdrawal	31	19

in the Low subgroup ranged in experience from less than 1 year to 6 years with a median of 3 years.

When the attitudes toward child behavior of the Lows are compared with those of the psychologists, 83 items were rated significantly differently by the two groups. Eighty of the 83 items, or 96%, were rated normal more often by psychologists than by teachers with relatively little experience. There was disagreement primarily in regard to the significance of aggressive behavior (57% of the items in that category were rated differently), regressive behavior (33% of the items here were differently rated), and affect expressions (31% of the items were judged differently). There was no difference in the ratings of the 18 phobic items, and relatively little disparity in judgments for inefficient behavior (16%) and fantasy-withdrawal behavior (16%). Communication problems and physical-psychosomatic disturbances produced only moderate disagreements (19% and 21%).

A similar chi-square analysis was done comparing the attitudes of teachers high in experience with psychologists on each of the scale items. In

TABLE 2

DEGREE OF DISAGREEMENT BETWEEN
HIGHLY EXPERIENCED AND RELATIVELY
INEXPERIENCED TEACHERS IN
SPECIFIC SCALE CATEGORIES

Category	Items in scale	Percentage of items disagreed upon
Physical-psychosomatic	71	28
Phobic	18	67
Aggressive	56	32
Affect	58	28
Communication	21	48
Regressive	15	13
Inefficiency	25	40
Fantasy-withdrawal	31	26

this comparison only 45 behavioral descriptions significantly differentiated the groups. Moreover, the patterns of disagreements between highly experienced teachers and psychologists, on the one hand, and less experienced teachers and psychologists, on the other hand, is very different. For one thing, the more experienced teachers did not nearly as often differ from psychologists in the direction of ascribing abnormality to a description of child behavior as did the less experienced teachers. As a matter of fact, the differences between the more experienced teachers and psychologists were likely to be as often in the direction of teachers considering the behavior to be benign when psychologists regarded it as being pathological as it was to be considered pathological when the psychologists rated it as being normal. (Only 52% of the differentiating items were rated normal by psychologists more often than by highly experienced teachers). Second, the areas in which the differences become manifest for the highly experienced teachers is very different from the Lows. More specifically, the Highs do not differ from the psychol-

ogists particularly in regard to aggressive, regressive, and affect behavior as do the Lows. The least disparity (0%), as a matter of fact, occurs in relation to regressive behavior; the greatest discrepancy (28% of the items in the category), occurs with the ratings of phobic behavior.

Finally, chi-square analyses of the item ratings for teachers high and teachers low in experience yielded the largest degree of discrepancy of all comparisons. Ninety-six, or 32.5% of the total number of behavioral descriptions, were rated significantly differently by these two subgroups. Interestingly, all 96 critical items were perceived to be normal more often by the highly experienced teachers as compared with the less experienced teachers.

Table 2 presents the percentage of items within each of the scale categories rated differently by highly experienced and relatively inexperienced teaching personnel. It will be noted that phobic behavior tends most often to be viewed differently by teachers of varying degrees of experience, and that there is considerable disagreement about behavior involving communication facility and efficiency.

Illustrative of the specific differences in ratings between the more experienced and the relatively less experienced teachers are the following items, all of which were regarded to be normal by the more experienced teachers and abnormal by the less experienced teachers:

Cries or whimpers
Plays with or fingers his mouth
Headache
At the slightest upset, coordination becomes poor
Is frightened in crowds
Is afraid of being alone in a wide open space
Child's thoughts are hard to understand
Lying

Lewis (1965), in reviewing the literature bearing on the "Continuity hypothesis," which states that "... emotional disturbance in a child is symptomatic of a continuing psychological process that may lead to adult mental illness [p. 465]," concluded that the acting-out child is more likely to become seriously disturbed as an adult than the timid, withdrawn child. He suggested that perhaps the judgments of teachers, as derived from the Wickman (1928) study, represented a more accurate appraisal of the pathology of children than the evaluations of clinicians, at least when adult psychiatric status is taken as the criterion. Irregardless of the validity of the perceptions of each of these groups, the study of the nature of the attitudes remains an important research problem since attitudes will influence markedly the interactions between the child and his teachers.

Beilin (1959) pointed out cogently that the attitudinal patterns of teachers and clinicians toward adjustment difficulties reflect in part their different roles, and that their roles, in turn, "influence the organization of their respective experiences [p. 22]." Since Beilin regards teachers to be essentially task-oriented, that is, concerned with the imparting of information and skills, and since mental health professionals are more concerned with preventing poor adjustment and promoting good adjustment, it is not surprising that these two groups will continue to perceive child behavior differently.

The present findings suggest that psychologists tend to be more accepting, or at least more tolerant, of a greater variety of child behavior than teachers, and tend to regard a wider range of behavior as being normal. Teachers, especially those who are relatively inexperienced, label much

more behavior as being abnormal. Teachers are especially critical of categories of behavior that may be referred to as aggressive, regressive, and emotionally expressive. The fact that the greatest degree of disagreement is found between experienced and inexperienced teachers reinforces the impression that actual exposure to child behavior is an important determinant of attitudes toward pathology.

The present study also bears on the frequently voiced criticism of clinicians as being overly sensitive to the pathological aspects of others and not sufficiently sensitive to their assets. The findings indicate that this criticism is probably unjustified since the clinicians were in fact much less prone to interpret behavior as being abnormal than the teachers.

Brief reference should be made to several methodological limitations. First, a number of teachers and psychologists who were given the Staten Island Scale either did not complete the form or failed to follow instructions and had to be eliminated for that reason. Thus, of the original sample of 145 teachers, only 118 could be employed for the analysis. Whether the respondents who cooperated differ in any essential respect from those who did not is not known. Second, although some precautions were taken against the respondents being influenced by others in making their ratings, the possibility still remains that some judgments were not made entirely independently.

Perhaps a more important problem is related to the ambiguous instructions provided the subjects. Many respondents found the task to be extremely difficult. A number took great pains to comment that since what is considered normal and abnormal is so closely related to the age level of the child, they could not arrive at a de-

cision. Moreover, some stated that since the degree, severity, frequency, nature of onset, duration, and circumstances surrounding the appearance of the symptom remained unspecified, their confidence level in arriving at a decision was extremely low. Nevertheless, it should be noted that since both the professionals and the teachers were faced with the same need to impose structure on the scale items, there is little likelihood that the ratings reflect systematic response biases that differ for the two groups.

It is suggested that the question of whether anchoring the concept of normality versus abnormality to specific age levels affects the ratings of groups of experts and teachers merits further research attention. Also, it might be possible to investigate the effect of increased structure in the description of each item, in terms of such characteristics as frequency of the symptom, on the judgments. Other extensions of this project would concern themselves with the variance contributed to teacher ratings of such variables as

their age, teaching competence, and socioeconomic status.

REFERENCES

- BEILIN, H. Teachers' and clinicians' attitudes toward the behavior problems of children: A reappraisal. *Child Development*, 1959, **30**, 9-25.
- GRIFFITHS, W. *Behavioral difficulties of children as perceived and judged by parents, teachers and children themselves*. Minneapolis: University of Minnesota Press, 1952.
- LEWIS, W. W. Continuity and intervention in emotional disturbance: A review. *Exceptional Children*, 1965, **31**, 465-475.
- MANDELL, W., & SILBERSTEIN, R. M. Children's psychopathology behavior rating scale. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, April 1965.
- SCHRUPP, M. H., & GJERDE, C. M. Teacher growth in attitudes toward behavior problems of children. *Journal of Educational Psychology*, 1953, **44**, 203-214.
- STOFFER, G. A. W., JR. Behavior problems of children as viewed by teachers and mental hygienists. *Mental Hygiene*, 1952, **36**, 271-285.
- WICKMAN, E. K. *Children's behavior and teachers' attitudes*. New York: Commonwealth Fund, 1928.

(Received July 6, 1966)

ERRATUM

Because of a statistical error, the article by Gerald W. Faust and Richard C. Anderson entitled "Effects of Incidental Material in a Programmed Russian Vocabulary Lesson" which appeared in the *Journal of Educational Psychology*, 1967, **58**, 3-10, incorrectly stated that in both experiments the Context groups recalled significantly more overtly-practiced Russian words than did No-Context groups. In fact, the overall comparison of Context and No-Context groups was not significant in either experiment. However, the Context groups did show significantly greater ($p < .05$) recall than the No-Context groups in both experiments for overtly-practiced Russian words which were easy to pronounce and for subjects with faster-than-median training times. Consequently, despite the statistical error, the main conclusions reported in the paper are supported by the data.

EFFECT OF FEEDBACK FROM TEACHERS TO PRINCIPALS¹

ROBERT W. DAW² AND N. L. GAGE

Stanford University

Each member of an experimental group of 151 elementary school principals was given feedback concerning his teachers' ratings of their actual and ideal principals on 12 behaviors. These principals were subsequently found to differ significantly, in the direction of teachers' preferences, from 143 principals in a control group. Initial differences in ratings were controlled by analysis of covariance. A 2nd, nonpretested control group did not differ from the pretested control group; hence the pretest itself did not produce the effect and difference between experimental and control groups was attributable to the feedback itself. 2 intervals between feedback and 2nd rating, 2 forms of feedback, the principal's age and experience, and the sequence and direction of the rating-scale items were found to be nonsignificantly related to the effect of the feedback. The results suggest that feedback of this kind improves the behavior of elementary school principals.

It is highly plausible that feedback regarding how others feel about one's behavior will affect one's behavior. Whether this maxim will hold under a given set of practical circumstances must, however, be determined empirically. In the present experiment, elementary school principals were told how their teachers rated them and an ideal principal; other principals, similarly rated, were not given this information.

One theoretical justification for hypothesizing that feedback of this kind

changes behavior has been developed by Gage, Runkel, and Chatterjee (1960). In brief, the rationale is that the feedback will inform some principals that their teachers evaluate their behavior less favorably than the principals might desire. If we assume that principals respect their teachers' opinions, we can expect this information to create in the principals a condition of imbalance (Heider, 1958), asymmetry (Newcomb, 1959), incongruity (Osgood & Tannenbaum, 1955), or dissonance (Festinger, 1957). To remove or reduce this condition, that is, to restore a condition of equilibrium or consistency, the principals are likely to change the behaviors concerned in the directions desired by the teachers. After enough time has elapsed to allow such behavior changes to occur and to be perceived by the teachers, a second description of the principals' behaviors by their teachers will reflect such changes.³

³ William McGuire (personal communication, August 15, 1966) has suggested that self-esteem theory—that is, that persons behave so as to maximize self-esteem, not to minimize inconsistencies, or discrepancies—is more relevant to our experiment. This

¹ A more detailed presentation of the data and instruments is available in the first author's doctoral dissertation, "Changing the Behavior of Elementary School Principals Through the Use of Feedback," on file in the Stanford University Library. The dissertation was written under the direction of the second author. Support for the research was provided by the Graduate Division of Stanford University and by a small grant (MHI544-01) from the United States Public Health Service. The present report was written by the second author during tenure as a Fellow at the Center for Advanced Study in the Behavioral Sciences and as a special Fellow of the National Institute of Mental Health.

² Now at Santa Maria Joint Union High School District, Santa Maria, California.

Positive results in experiments on the effect of feedback of ratings have previously been obtained (Bryan, 1963; Gage, 1963; Gage, Runkel, & Chatterjee, 1963). In those experiments, teachers were rated by their pupils and then, on a subsequent rating by the same subjects, were found to differ significantly, and in the direction of the raters' ideals, from a control group not given the ratings.

The present experiment was aimed at determining whether the same effects would be found with feedback from teachers to principals. It also incorporated some refinements in design that permitted testing rival hypotheses as to the cause of the change in rated behavior. In addition, data were gathered concerning different time intervals, forms of feedback, and personal characteristics of the principals.

METHOD

Subjects

The subjects were 455 elementary school principals in all the counties in California in the fall of 1962. Because they were assigned at random, the principals in the experimental and control groups had about the same number of pupils and teachers, and were similar in sex, age, educational level, and years of experience as a principal, as shown in Table 1. Besides the experimental (E) group, there were two control groups: Control Group C₁, which was rated on both the first and second occasions, and "posttest only" Control Group C₂, which was rated only on the second occasion.

Each superintendent of an elementary school district (including unified districts) in California with more than one principal was invited to send the name of the first principal in the alphabetical listing of principals in his district, and then that principal was invited to participate. In districts with

kind of alternative to consistency theory has been outlined by Deutsch, Krauss, and Rosenau (1962), Steiner and Rogers (1963), and McGuire and Millman (1965).

TABLE 1
DISTRIBUTIONS, MEANS, AND RANGES ON
VARIOUS VARIABLES FOR EXPERIMENTAL
AND CONTROL GROUPS

Variable	Experimental group (E)	Control group 1 (C ₁)	Control group 2 (C ₂)
Kind of district			
District with superintendent other than principal	105	111	115
District where the principal also acts as the superintendent	46	32	46
Sex			
Male	137	130	146
Female	14	13	15
Education			
No B.A. degree	1	0	1
B.A. degree	49	40	44
M.A. degree	97	100	115
Ed.D. degree	3	3	1
Ph.D. degree	1	0	0
Number of students in the school			
M	509	512	507
Range	185-1011	185-1157	200-1175
Age			
M	42	43	41
Range	27-63	20-63	27-63
Number of years of experience as a principal			
M	8.3	8.6	7.9
Range	1-36	1-32	1-35
Number of teachers in the school			
M	14.8	15.3	15.5
Range	8-34	8-37	8-39

Note.—For the Experimental Group, *N* = 151; for Control Group 1, *N* = 143; for Control Group 2, *N* = 161.

a single principal who also acted as superintendent, the superintendent-principal was directly invited to participate. Although initial attrition was about 25%, the subsequent rate of participation was never less than 93% of the principals invited. That is, of the 1007 original contacts, 752 yielded the name of the person who participated. To insure the anonymity of the teacher responses, all schools with less than eight teachers were eliminated, leaving 500 principals. Of these, 455 completely met all other requirements for inclusion in the experiment. Obviously, this final group may have been biased toward containing superintendent-principals (a) interested in what their teachers thought of their actions or (b) reluctant to refuse to participate.

Methods and Schedule of Data Collection

The study was made in the school year 1962-63. Letters of invitation were sent on November 30. To minimize the possibility that principals would discuss the project

with one another, only one principal was invited from each district. By January 5, 752 principals' names had been received; of these, 252 were excluded because their schools had less than eight teachers. Between January 2 and 9, booklets entitled "What Do They Expect?" (WDTE) were mailed to 340 principals in Groups E and C₁. These booklets, presenting the research as the tryout of a "Principal's Information Project," required the principal to rate himself and his ideal principal on 12 items. In addition, the principal was asked to provide the information summarized in Table 1 and to indicate how many Teacher Opinion Booklets would be needed to collect ratings of the principal from his teachers. (The WDTE booklet was mailed to Group C₂ just prior to the mailing of the posttest materials.) Principals were randomly assigned to Groups E, C₁, and C₂ prior to the mailing of the WDTE booklet. The rate of return of these booklets for the three groups ranged from 94.0 to 98.7%.

Teacher Opinion Booklets were mailed to the principals on January 21; only eight of the 336 booklets mailed were not returned. In these booklets, the teachers rated their actual and ideal ("best imaginable") principal on the same 12 items.

On February 11, booklets entitled "Report on Your Teachers' Opinions" (RYTO) were mailed to the principals in Group E. In these booklets, as shown in Figure 1, the

principal was given, for each of the 12 items, histograms showing the percentages of his teachers who described him and their ideal principal with each of the six response-alternatives. A randomly chosen half of the experimental group (the "a + h" group) received both the histograms and the medians (indicated by arrows) of the ratings of the actual and ideal principal; the remaining principals (the "a" group) were given only the medians of the ratings by their teachers. These RYTO booklets were withheld from Groups C₁ and C₂ until after the second round of ratings of principals by their teachers.

A follow-up questionnaire, designed to encourage careful reading of the RYTO's, was mailed on February 25 to all principals in Group E. This questionnaire and a reminder elicited responses concerning the RYTO from 93% of the experimental group.

The second round of Teacher Opinion Booklets was mailed to a randomly chosen half of Groups E and C₁ on March 25. A letter had been mailed a week in advance asking the principal to set aside a specific day for the administration of these instruments. They were returned with little delay. Teacher Opinion Booklets were mailed to the second half of Groups E and C₁ on May 6. The use of two posttest dates permitted determining whether the effect of the feedback changed with the length of interval between feedback and posttest.

Item: Acts Promptly to Fulfill Teacher Requests

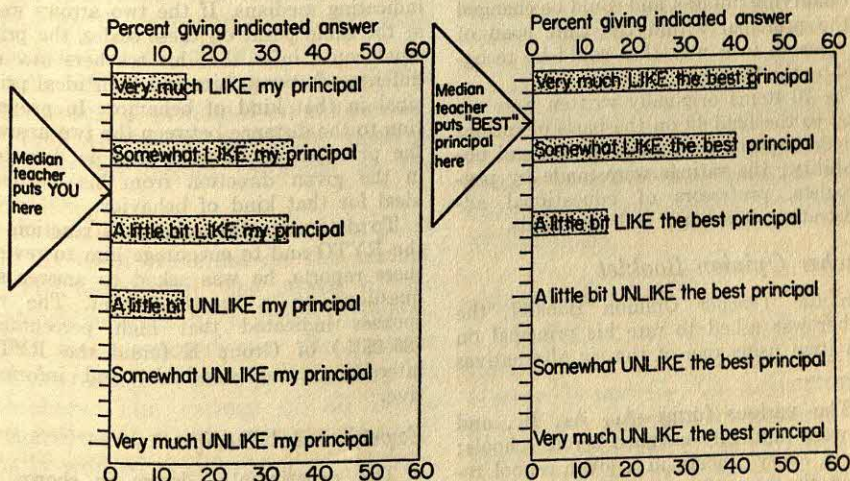


FIG. 1. Form of feedback, arrow-plus-histogram, in "Report on Your Teachers' Opinions."

Items Concerning Principal Behavior

The teachers described their actual principal and their ideal principal on the following 12 items concerning principal behavior:

1. Encourages teachers with a friendly remark or smile.
2. Gives enough credit to teachers for their contributions.
3. Does not force opinions on teachers.
4. Enforces rules consistently.
5. Criticizes without disparaging the efforts of teachers.
6. Informs teachers of decisions or actions which affect their work.
7. Gives concrete suggestions for improving classroom instruction.
8. Enlists sufficient participation by teachers in making decisions.
9. Demonstrates interest in pupil progress.
10. Interrupts the classroom infrequently.
11. Displays much interest in teachers' ideas.
12. Acts promptly in fulfilling teacher requests.

The items were based on ideas obtained from Campbell and Gregg (1957), Gross, Mason, and McEachern (1958), Guba and Bidwell (1958), Medsker (1956), and various elementary school teachers. Each item was worded in both positive and negative directions in Forms A and B, respectively. Forms A₁ and B₁ placed the items in reverse order from that in Forms A₂ and B₂.

The items were intended to deal with behaviors that could be expected to occur frequently, could be briefly described without qualifying phrases, and could be changed by the principal within the time span of the research in a way that could be recognized by the teachers.

The 70 items originally written were reduced to the final 12 on the basis of ratings of their importance, improvability, and noticeability; the ratings were made by psychologists, professors of educational administration, teachers, and principals.

Teacher Opinion Booklet

In the Teacher Opinion Booklet, the teacher was asked to rate his principal on each item using one of the six alternatives

'The various forms—A₁, A₂, B₁, and B₂—were randomly assigned across schools; that is, every teacher in a given school received the same form.

shown in Figure 1. Then he was asked to rate his ideal principal on the same 12 items. The teachers read the directions silently while the principal read them aloud. The teacher wrote his responses on a card, put the card in an envelope, and sealed the envelope. On the cover of his booklet, the teacher read: "Your answers will be sealed in an envelope by you and sent directly to Stanford University. No one at your school or in your district will know how you answered these questions." Further to insure the teachers' privacy, the principal was directed to stand far enough away from his teachers to prevent him from seeing their papers, to make certain that all teachers' answer cards were sealed in the envelope, to require the teachers to put their sealed envelopes into the large mailing envelope, to permit his teachers to see him moisten and close the clasp of the large envelope, and finally to ask one of his teachers to drop the envelope in a United States mail box. The administration of the questionnaires was standardized as fully as possible with printed directions to the principals and teachers so as to assure both groups that their anonymity would be preserved. It is noteworthy that all answer cards came back in sealed envelopes.

"Report on Your Teachers' Opinions"

The "Report on Your Teachers' Opinions" consisted of a booklet containing two charts for each of the 12 items, as shown in Figure 1. The discrepancy between the teachers' descriptions was implicit in the vertical distance between the two arrows indicating medians. If the two arrows were at the same point on these scales, the principal could infer that his teachers saw no difference between him and their ideal principal in that kind of behavior. In proportion to the distance between the two arrows, the principal could infer that he departed in the given direction from his teachers' ideal for that kind of behavior.

To determine the principal's reaction to the RYTO and to encourage him to review these reports, he was asked to answer six questions on a reaction sheet. The responses indicated that high percentages (85-95%) of Group E found the RYTO interesting, understandable, and informative.

Experimental Design

The experimental design is shown in Figure 2. Here, X represents the experi-

Group	Pretest	Treatment	6 weeks	12 weeks
E a-short	O ₁	X _a	O ₂	
E a-long	O ₃	X _a		O ₄
E a+h-short	O ₅	X _{a+h}	O ₆	
E a+h-long	O ₇	X _{a+h}		O ₈
C 1-short	O ₉		O ₁₀	
C 1-long	O ₁₁			O ₁₂
C 2-short			O ₁₃	
C 2-long				O ₁₄

FIG. 2. The experimental design.

mental treatment (feedback), O refers to the process of observation (ratings), Xs and Os in a given row are applied to the same subjects, the left-to-right dimension indicates temporal order, and the rows represent equivalent samples of persons. X_a represents "arrow-only" feedback; X_{a+h} represents "arrow-plus-histogram" feedback. Group C₂ (posttest only) was used to allow comparisons free of any effect attributable to unintended feedback received by Group C₁ through participating in the pretest; these comparisons would show whether the pretest itself produced changes in behavior.

RESULTS

In describing the results, we shall refer to the protocols obtained from the teachers as follows:

Pre-ACT—the teacher's description of his *actual* principal on the pretest.

Post-ACT—the teacher's description of his *actual* principal on the posttest.

Pre-IDL—the teachers' description of his *ideal* principal on the pretest.

For each item, the mean of the ratings received by a group of principals was computed over the medians of the ratings of each principal by his teachers. The ratings of all items, regardless of whether they were originally worded positively or negatively, were converted to a scale in which 1

signified the desirable end, and 6 the undesirable end, of the continuum.

Pre-ACT

The random assignment of principals to Groups E and C₁ should have made them equivalent. But the pre-ACT means of these groups differed at the .05 level on five of the 12 items and on the mean of the 12 items. Also, the direction of the difference was the same for all 12 items, namely, the direction favoring Group E. Presumably, however, the analysis of covariance, to be described below, eliminated this pretest bias to the extent that the pretest scores were reliable.

The difference between Groups E and C₁ in pre-ACT means might have resulted from a greater tendency of subjects receiving unfavorable feedback on pre-ACT to drop out of Group E. In that event, the remaining members of Group E would be those who had received more favorable pre-ACT ratings. But when this possibility was investigated, it was found that, in fact, the few drop-outs from Groups E and C₁ did not differ on pre-ACT in this way, and the difference remained unexplained.

The two experimental subgroups differing in the type of feedback provided (arrow-only and arrow-plus-histogram) did not differ significantly on any of the 12 pre-ACT means. Nor were there any significant differences in pre-ACT means among the subgroups given forms differing in the direction of wording or the sequence of the items.

Adjusted Post-ACT

The effect of the feedback was measured by the difference between Groups E and C₁. To adjust for the pre-ACT differences between the groups, analysis of covariance was used. The pre-ACT rating served as

TABLE 2

MEANS OF TEACHERS' RATINGS OF ACTUAL PRINCIPALS IN EXPERIMENTAL GROUP AND CONTROL GROUPS C₁ AND C₂

Item	Means							Difference between adjusted post-ACT Means	F
	pre-ACT		post-ACT			Adjusted post- ACT			
	E	C ₁	E	C ₁	C ₂	E	C ₁		
1	1.32	1.39	1.31	1.50	1.44	1.38	1.43	.05	11.6**
2	1.47	1.57	1.45	1.63	1.60	1.50	1.58	.08	5.0*
3	1.54	1.61	1.50	1.77	1.77	1.61	1.66	.05	15.5**
4	1.68	1.75	1.59	1.84	1.96	1.69	1.74	.05	13.5**
5	1.34	1.51	1.36	1.56	1.56	1.42	1.49	.08	6.7*
6	1.42	1.46	1.38	1.61	1.67	1.48	1.51	.03	17.0**
7	1.72	1.84	1.64	2.04	1.93	1.81	1.87	.06	25.1**
8	1.58	1.63	1.50	1.80	1.86	1.63	1.67	.04	19.0**
9	1.28	1.37	1.27	1.36	1.38	1.30	1.34	.04	2.2
10	1.19	1.26	1.20	1.29	1.50	1.23	1.26	.03	3.3
11	1.40	1.55	1.40	1.59	1.57	1.45	1.54	.09	4.4*
12	1.51	1.61	1.44	1.66	1.71	1.52	1.57	.05	8.7**
1-12	1.45	1.55	1.42	1.64	1.65	1.49	1.57	.08	19.4**

Note.—In this and all other tables, means refers to a scale in which a score of 1 was assigned to "Very much LIKE," rating scale alternative, 2 to "Somewhat LIKE," etc., to 6 for "Very much UNLIKE." For the Experimental Group (E), $N = 151$; for Control Group 1 (C₁), $N = 143$; for Control Group 2 (C₂), $N = 161$.

* $p < .05$.

** $p < .01$.

the control variable, the post-ACT rating as the dependent variable, and the feedback as the independent variable. Although the 12 items are correlated, it is considered worthwhile to examine results for each item individually as well as for the mean over all 12 items. Analyses of covariance for each of the 12 items and for the mean over Items 1-12 yielded the results shown in Table 2. The difference between the adjusted post-ACT means was significant at the .05 level or better for all but two of the 12 items; for the mean over Items 1-12, it was significant at the .01 level. In all cases, the difference between the adjusted post-ACT means favored the experimental group. Only on Items 9 and 10 did the adjusted post-ACT means not differ at even the .05 level of significance. But even on these items, the difference favored the experimental group. Item 9, "Demonstrates interest in pupil progress," may have been too difficult for the teachers

to perceive in the period allowed. Item 10, "Interrupts the classroom infrequently," may have been rated too favorably on the pretest to allow sufficient room for improvement; the pre-ACT means on Item 10 were the most favorable (had the lowest numerical value) of all 12 items.

Posttest-Only Control Group

The "posttest-only" control group (C₂) was used to eliminate certain possible attenuating effects on the comparisons. Such effects might have resulted from any feedback, or sensitization to the items of behavior, that might have been received by the pretested control group (C₁) as a result of their participation in the pretest. If such sensitization occurred, Group C₁ would differ less from Group E than would a nonpretested control group (C₂) which received neither the feedback nor the pretest. Here the comparisons must be made in terms of unadjusted posttest means, since there were no pretest means with which to adjust the posttest means of Group C₂.

Table 2 also shows the means for Group C₂. The means for Group C₁ and C₂ did not differ significantly. Both of these groups did differ, in the same direction, from Group E. Hence, the pretest in itself did not affect Group C₁, and the improvement in Group E must be ascribed to the feedback alone, not to the feedback plus the pretest.

Adjusted Post-ACT Minus Pre-IDL

The teachers' initial ratings of their ideal principal (pre-IDL) make possible an interpretation of the direction of the difference between adjusted post-ACT means. It should be recalled that the feedback informed Group E as to how their teachers rated the ideal principal. The feedback should influence the principal to

TABLE 3

ADJUSTED MEAN POST-ACT AND PRE-IDL RATINGS FOR THE EXPERIMENTAL GROUP AND CONTROL GROUP C_1

Item	Adjusted post-ACT M		pre-IDL M		Adjusted post-ACT minus pre-IDL	
	E	C_1	E	C_1	E	C_1
1	1.38	1.43	1.09	1.08	.29	.35
2	1.50	1.58	1.13	1.14	.37	.44
3	1.61	1.66	1.21	1.22	.40	.44
4	1.69	1.74	1.20	1.17	.49	.57
5	1.42	1.49	1.13	1.17	.29	.32
6	1.48	1.51	1.09	1.10	.39	.41
7	1.81	1.87	1.19	1.25	.62	.62
8	1.63	1.67	1.20	1.23	.43	.44
9	1.30	1.34	1.10	1.13	.20	.21
10	1.23	1.26	1.15	1.19	.08	.07
11	1.45	1.54	1.13	1.18	.32	.36
12	1.52	1.57	1.13	1.16	.39	.41
1-12	1.49	1.57	1.14	1.17	.35	.40

Note.—For the Experimental Group, $N = 151$; for Control Group C_1 , $N = 143$.

change in that direction. Hence, the difference between adjusted post-ACT and pre-IDL should be smaller for Group E than for Group C_1 . As Table 3 shows, most of these differences were indeed smaller for the experimental group. At the time of the posttest, the principals who received feedback came closer to their teachers' desires. Since the items are interdependent, they do not, of course, provide 12 separate tests of the overall hypothesis. But the means over all 12 items also differed in the expected direction.

Interval, Form of Feedback, and Other Variables

In the experiment by Gage, Runkel, and Chatterjee (1960), experimental group subjects changed more if they had a longer interval between feedback and posttest. But the time interval was relatively short, ranging from 18 to 59 days. In the present experiment, the experimental and control groups were assigned at random to either a

6-week or a 12-week interval group. Analysis of covariance revealed no significant difference in the two groups' adjusted post-ACT means. Further, there was no significant effect due to the Feedback $\times$ Interval interaction. In short, the effect of the feedback was not a function of the interval over which it was measured.

One randomly chosen half of the experimental groups received feedback in the form of median ratings (arrows) only, and the other half received median ratings plus the frequency distributions (histograms) of the ratings. Either kind might reasonably be predicted to be the more effective: The median-only might be sharper, less ambiguous; on the other hand, the median-plus-distribution might be more convincing. But no significant difference was found for any item between the adjusted post-ACT means of the subgroups receiving these two kinds of feedback.

When the principals were divided into subgroups on the basis of age (40 or younger versus 41 or older), experience (5 years or less as elementary school principal versus 6 years or more), and form (A_1 , A_2 , B_1 , B_2) of the Teachers' Opinion Booklet, analyses of covariance yielded no significant difference at the .05 level due to main effects or interactions. Thus, the significant results that occurred due to feedback did not seem to vary with age, experience, or form. The latter findings show that positional or directional sets did not significantly affect the teachers' responses.

DISCUSSION

All in all, the results indicate that the feedback affected changes in the principals' behavior. Various questions remain, however, for subsequent research. First, we must recall the unexplained pretest differences between

the experimental and control groups despite their random assignment. Second, we must ask whether methods of measuring behavior other than re-ratings by the same teachers would reveal the same kinds of behavior change. Would observations by trained observers produce confirming evidence? Would interviews of the principals reveal the process by which feedback operates, and show whether the principals were consciously attempting to change their behavior? Would disguised tests of teacher "morale" reflect the desirable changes in rated principal behavior?

Third, we should determine better whether the improvements in the post-ACT ratings reflect mere improved "halo effect" or actual changes in specific behaviors. One way to proceed on this question would be to collect post-ACT ratings on uncorrelated items not dealt with in the feedback; if behavior also improves on such items, the significance of the feedback on specific items must be questioned. Similarly, items differing widely in "changeability" could be compared as to the amount of change principals exhibit in them; presumably, if the ratings reflect more than general impressions, relatively unchangeable behaviors should be rated as changing less.

If research allays the skepticism implicit in these suggestions, further attention should be given to ways of enhancing the effectiveness of feedback. The behavior of teachers, principals, and persons in many similar roles could be made more effective by applying the results of such a program of research.

REFERENCES

BRYAN, R. C. *Reactions to teachers by students, parents and administrators*. (United States Office of Education, Cooperative Research Project No. 668.) Kalamazoo: Western Michigan University, 1963.

- CAMPBELL, R. F., & GREGG, R. T. (Eds.) *Administrative behavior in education*. New York: Harper, 1957.
- DEUTSCH, M., KRAUSS, R. M., & ROSENAU, N. Dissonance or defensiveness? *Journal of Personality*, 1962, **30**, 16-28.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.
- GAGE, N. L. A method for "improving" teacher behavior. *Journal of Teacher Education*, 1963, **14**, 261-266.
- GAGE, N. L., RUNKEL, P. J., & CHATTERJEE, B. B. *Equilibrium theory and behavior change: An experiment in feedback from pupils to teachers*. (Report No. 6 in Studies in the generality and behavioral correlates of social perception.) Urbana: Bureau of Educational Research, College of Education, University of Illinois, 1960.
- GAGE, N. L., RUNKEL, P. J., & CHATTERJEE, B. B. Changing teacher behavior through feedback from pupils: An application of equilibrium theory. In W. W. Charters & N. L. Gage (Eds.), *Readings in the social psychology of education*. Boston: Allyn & Bacon, 1963. Pp. 173-181.
- GROSS, N., MASON, W., & McEACHERN, A. W. *Explorations in role analysis: Studies of the school superintendency role*. New York: Wiley, 1958.
- GUBA, E. G., & BIDWELL, C. E. *Administrative relationships*. Chicago: Midwest Administration Center, University of Chicago, 1958.
- HEIDER, F. *The psychology of interpersonal relationships*. New York: Wiley, 1958.
- MCGUIRE, W. J., & MILLMAN, S. Anticipatory belief lowering following forewarning of a persuasive attack. *Journal of Personality and Social Psychology*, 1965, **2**, 471-479.
- MEDSKER, L. L. The job of the elementary school principal as viewed by teachers. Unpublished doctoral dissertation, Stanford University, 1956.
- NEWCOMB, T. M. Individual systems of orientation. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 3. New York: McGraw-Hill, 1959. Pp. 384-422.
- OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review*, 1955, **62**, 42-55.
- STEINER, I. D., & ROGERS, E. D. Alternative responses to dissonance. *Journal of Abnormal and Social Psychology*, 1963, **66**, 128-136.

(Received July 11, 1966)

New from Addison Wesley

INTERACTION ANALYSIS: THEORY, RESEARCH, AND APPLICATION

By EDMUND J. AMIDON, *Temple University*, JOHN B. HOUGH, *Syracuse University*, With an Introduction by NED FLANDERS, *University of Michigan*

The purpose of this supplementary text is to bring together under one cover a representative selection of papers dealing with the subject of Interaction Analysis to those involved in the training of teachers. The book is divided into three parts as follows: I) the educational and social psychological theories of classroom climate, II) the development and empirical development of interaction analysis as a technique for assessing classroom social-emotional climate, III) the application of interaction analysis to problems of teacher-training.

SPECIAL FEATURES

1. The concepts of theory, research and application in that order are used as organizing themes.
2. The book contains many papers not readily available through other sources.
3. The primary source papers deal with a contemporary and highly effective means for analyzing classroom verbal interaction.

In Press.

eeing, Knowing and Believing

By JONAS F. SOLTIS, *Columbia University*.

This complete and "readable" treatment of the language of visual perception in contemporary philosophy is designed to provide a theoretical framework which will permit more precise future discussions of the phenomenon of visual perception by philosophers and cognitive psychologists.

The book fully examines such ideas as "recognition," "identification," "illusion," and the more basic notion of "seeing" itself.

156 pp. \$3.95

Assessing Behavior: Readings in Educational and Psychological Measurement

By JOHN T. FLYNN and HERBERT GARBER, *University of Connecticut*.

Specifically designed for senior, graduate and advanced graduate courses in tests and measurement, this book includes both a psychometric and a clinical approach, giving equal emphasis to clinical and experimental positions regarding behavior assessment. The book's main purpose is to provide a wider, deeper acquaintance with the history, problems, and strategies of psychological measurement.

Paperbound 377 pp. \$3.95

Addison-Wesley
PUBLISHING COMPANY, INC.

Reading, Massachusetts 01867



*A new revision
of major significance . . .*

APA PUBLICATION MANUAL 1967 Revision

The APA PUBLICATION MANUAL has undergone a complete revision for the first time in 10 years. Although the revised edition prescribes the same basic format for manuscripts as the previous edition, major changes and additions were made in the text in an effort to clarify and simplify the instructions on how manuscripts should be prepared.

The APA PUBLICATION MANUAL is intended primarily as a guide for authors and editors in the preparation of material to be published in APA journals. However, in preparing the original edition it was recognized that the MANUAL could, and should, be written to accommodate a wider range of manuscripts. The revised PUBLICATION MANUAL has retained that original concept.

April, 1967

\$1.50

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D. C. 20036

RUSSIAN MONOGRAPHS ON BRAIN AND BEHAVIOR

A series of translations of Russian Monographs concentrating upon the historical background of investigations in neurophysiology and conditioned behavior as well as upon certain contemporary advances in these fields.

The translation program was initiated by H. W. Magoun and D. B. Lindsley and has been continued by the latter. The complete series will consist of eleven monographs published in hard-cover book form. To date, three of the eleven monographs have been printed and are available through the American Psychological Association.

Available now . . .

ESSAYS ON THE HISTORY OF PHYSIOLOGY IN RUSSIA

By Kh. S. Koshtoyants.

These essays review broadly the history of the development of physiology as a science in Russia, beginning in the early 18th century and continuing to modern times. They are more than a history for they deal with the personal characteristics and strivings of great scientists who forged teaching and research disciplines in anatomy, physiology, and psychology.

321 pp. \$4.00

AUTOBIOGRAPHICAL NOTES

By I. M. Sechenov.

Sechenov, revered as the father of physiology in Russia, was born in 1829. His scientific and academic pursuits, though always stimulating and challenging, were fraught with difficulties due to political interference and harassment. These **Autobiographical Notes**, written about a year before his death in November, 1905, provide a vivid picture of his history.

174 pp. \$4.00

ORIENTING REFLEX AND EXPLORATORY BEHAVIOR

Edited by L. G. Voronin, A. N. Leontiev, A. R. Luria, E. N. Sokolov, and D. S. Vinogradova.

A collection of papers read at a Conference organized in February, 1957 by the Institute of Defectology of the Academy of Pedagogical Sciences, RSFSR, and by the Department of Psychology and Higher Nervous Activity of the M. V. Lomonosov Moscow State University. As the varied experiments and points of view in this collection of papers will reveal, a serious attempt has been made to establish the role and significance of the orienting-exploratory reaction, not only in the elementary behavior of lower animal forms, but also in the complex behavior of higher forms, including man and his cognitive functions.

462 pages \$4.00

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1200 Seventeenth St., N. W., Washington, D. C., 20036

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Diary No. 93
10 9 70
File No
Lib
Larsen, H. J. Research
Number 4

Volume 58

August 1967

CONTENTS

- Effect of Similarity in Structure, Meaning, and Sound on Paired-Associate Learning*..... 189
LOWELL SCHOER
- Second-Try Recall, Recognition, and the Memory-Monitoring Process*..... 193
JOSEPH T. HART
- The Many Faces of Intelligence*..... 198
CHARLES E. WERTS
- The Many Faces of Talent: A Reply to Werts*..... 205
JOHN L. HOLLAND AND JAMES M. RICHARDS, JR.
- Student Perceptions of College Persistence, and Their Intellectual, Personality, and Performance Correlates*..... 210
EDMOND MARKS
- Learning of Logical Connectives by Adolescents with Single and Multiple Instances*..... 222
JAMES YOUNISS AND HANS G. FURTH
- Control of Hyperactivity by Social Reinforcement of Attending Behavior*..... 231
K. EILEEN ALLEN, LYDIA B. HENKE, FLORENCE R. HARRIS, DONALD M. BAER, AND NANCY J. REYNOLDS
- School-Related Attitudes of Culturally Disadvantaged Elementary School Children*..... 238
DANIEL C. NEALE AND JOHN M. PROSHEK
- Acquisition of Problem-Solving Strategies in Young Children and Its Relation to Verbalization*..... 245
CAROLYN STERN
- Acquisition and Retention of Discrimination Learning Sets in Lower-Class Preschool Children*..... 253
PHYLLIS A. KATZ

EFFECT OF SIMILARITY IN STRUCTURE, MEANING, AND SOUND ON PAIRED-ASSOCIATE LEARNING

LOWELL SCHOER

University of Iowa

The purpose of the study was to determine the effect of 3 types of stimulus similarity on paired-associate learning of 5th-grade Ss. 80 Ss learned each of 4 lists (total $N = 320$). The stimuli in List I were similar in meaning, those in List II in sound, and those in List III in structure. List IV was the control list. List III performance was significantly lower than that on the other 3 lists. None of the other list differences were significant. There was also a significant ability effect, but the interaction between lists and ability level was not significant.

The relationship between stimulus similarity and learning has been found to be consistent with predictions made from Gibson's (1940) theory when forms and nonsense syllables have been used as stimuli (e.g., Feldman & Underwood, 1957; Gibson, 1942). The results have not been nearly so consistent when adjectives have been used (Beecroft, 1956; Underwood, 1953). This might suggest that Gibson's is an adequate theory of "form learning" but not of "word learning" (Underwood, 1961), or it might suggest that it is an adequate theory of word learning only when the dimension of similarity involved is that which might be associated with primary rather than secondary generalization. The primary purpose of the present study was to investigate this latter possibility.

Basic to such an investigation is some statement of the difference between primary and secondary generalization. For purposes of the present study, primary or sensory stimulus generalization is defined as that generalization which occurs as a function of the physical similarity among the stimuli. Secondary stimulus generalization is, on the other hand, defined as that generalization which occurs as a function of factors other than physical similarity. By these defini-

tions the generalization that would occur as a result of using the words "age" and "ago" as stimuli would be primary generalization, while that which would occur as a result of using "large" and "big" as stimuli would be secondary generalization. In addition to structure and meaning, words may also be similar in sound, for example, "ate" and "eight." Such similarity, because it is a dimension of similarity that results from the sounds of the words when they are spoken rather than similarity in the structure of the stimuli, would seem to be more closely associated with secondary than with primary generalization.

Basically, the present study was designed to determine the effect of stimulus similarity in structure, meaning, and sound on paired-associate learning as a means of investigating the operation of primary and secondary generalization in word-learning.

Because there is some suggestion from the literature that high- and low-ability learners differ in their susceptibility to the effects of generalization (Schoer, 1963), the study was also designed to investigate the effects of these three sources of generalization on learners of varying abilities.

METHOD

Lists

As indicated above, words may be similar in meaning, sound, or structure and a pair of words may be similar in one of these dimensions and dissimilar in the other two. The first step in setting up the learning lists employed in the present study involved choosing three pairs of words similar in meaning but dissimilar in sound and structure, three pairs of words similar in sound, but dissimilar in meaning and structure, and three pairs similar in structure, but dissimilar in meaning and sound. All the words used were chosen from among the first 2,000 in the Thorndike-Lorge word list (Thorndike & Lorge, 1944). The six words of each type constituted the stimulus words for one learning list. The response words were the same for all three lists and consisted of six words dissimilar in meaning, sound, and structure. A fourth list, the control list, used the same six response words as the three lists described and six stimulus words which were also dissimilar in meaning, sound, and structure. The four learning lists are given below.

I (Meaning)	II (Sound)
TALK-from	I-drop
SPEAK-kept	EYE-move
LARGE-move	WON-care
BIG-drop	ONE-poor
GLAD-care	ATE-from
HAPPY-poor	EIGHT-depth
III (Structure)	IV (Control)
AGE-move	SPEND-move
AGO-drop	HAPPEN-poor
OWN-from	GOLD-care
ONE-kept	BIT-kept
ELSE-poor	TALL-from
EAST-care	LAUGH-drop

Subjects

The subjects (Ss) were students in 16 fifth-grade classrooms in the Davenport, Iowa, public schools. Each of the four lists was learned in four classrooms.

The Iowa Tests of Basic Skills composite was used as the measure of ability and the levels used were those set off by the quartiles on the Iowa norms. Because the statistical design employed required proportionality, a number of Ss who were in the classrooms where the lists were learned were not in-

cluded in the analysis. The final number in the analysis was 320, with 20 Ss from each level on the ITBS for each of the four learning lists.

Apparatus and Procedure

The learning materials were presented to Ss in intact classes using a Dunning Animatic to project the lists onto a screen in the front of the room. All the classrooms were equipped with pull-down screens and the Dunning Animatic is completely portable so a minimum of time was required to set up the needed equipment in the classrooms. After the equipment was set up, all Ss were given a sheet of paper and asked to spell the six response words. Although this involved some response pretraining it was felt to be necessary because Ss were going to write their responses in the experiment, which necessitated finding out if there were any students who could not spell them. All such students, and there were very few of them, were dropped from the sample.

After Ss had spelled the response words, the nature of the experiment and the use of the answer sheet was explained to them. They were told that they would be shown pairs of words and that they should study the pairs so that later when they were shown only the first word of the pair they could write down the second word. The answer sheet they were given provided space to write their responses. These answer sheets were set up in such a fashion that Ss could, after each response trial, quickly fold their answers on that trial under and out of sight. In addition, each S was provided with a mask to be used to cover the responses after he made them on a given trial. In this way all previous responses given by S were out of sight when he wrote the response to a given stimulus word.

A practice list of three pairs of words was used to be sure Ss understood the directions and to give them practice in using the answer sheet. The fifth-grade Ss used had little difficulty understanding what they were to do and became relatively proficient in the use of their answer sheets during the period of practice provided by the practice list.

Learning and recall trials were alternated during the learning of both the practice and the experimental lists. During the learning trials each pair was presented for 2 seconds. After all the pairs had been presented there was a 5-second delay interval after which only the stimulus words of each pair were presented.

The *S* was given 5 seconds to write down the word that "went with" each stimulus word as it was shown. This recall trial was followed by a second learning trial which was in turn followed by a second recall trial, etc. The *Ss* were given four learning and four recall trials on the practice list and six learning and six recall trials on the experimental list. Each *S* learned only one list.

RESULTS

The criterion measure for the analysis was the number of correct responses on the sixth recall trial. The data were analyzed using a factorial design (Lindquist, 1953).

The means for the four ability levels on the four lists are shown in Table 1. The *F* for lists was significant at well beyond the .01 level ($F = 636$, $df = 3/304$) as was the *F* for ability ($F = 5.61$, $df = 3/304$). The interaction between lists and ability was not significant ($F = 1.15$, $df = 9/304$).

Lindquist (1953, p. 214) describes a procedure for making individual comparisons of row or column means. By this procedure, a difference of .61 was found to be necessary for significance at the .05 level. Because the *N* for rows is the same as the *N* for columns this is the critical difference for significance at the .05 level for both rows and columns. Applying the critical difference .6 to the differences among the lists shows List III performance to be significantly lower

TABLE 1

LIST MEANS BY ABILITY LEVEL ON SIXTH RECALL TRIAL

Quarters on Iowa Tests of Basic Skills	List				
	I	II	III	IV	Total
4	4.7	4.9	3.2	5.0	4.5
3	4.8	4.5	3.3	3.6	4.1
2	3.9	3.2	3.3	4.3	3.7
1	3.4	3.8	2.4	3.4	3.3
Total	4.2	4.1	3.1	4.1	

Note.—*N* = 20 per cell.

TABLE 2

NUMBER OF SUBJECTS WHO GAVE RESPONSES THAT WERE INCORRECT FOR STIMULUS GIVEN BUT CORRECT FOR OTHER SIMILAR STIMULUS IN LIST ON TRIAL 6

Subjects	List			
	I	II	III	IV
Did	18	11	29	8
Did not	62	69	51	72

Note.— $\chi^2 = 19.94$; $df = 3$; $p < .01$.

than the performance on any of the other lists. None of the other list differences even approaches significance. Applying the same critical difference to the ability levels shows the performance of *Ss* in the upper quarter to be significantly higher than that of those in the second and bottom quarters, while that of *Ss* in the third quarter is significantly higher than that of *Ss* in the bottom quarter.

The list means for all sex recall trials combined were 3.5, 3.3, 2.9, and 3.4 for Lists I, II, III, and IV respectively.

There is, of course, the possibility that these list differences are due to some factor other than similarity. Support for the similarity explanation could be provided by an analysis of the errors on Trial 6, particularly errors which involved *S* giving a response which was incorrect for the stimulus shown but correct for the other stimulus of the pair in the list. The number of *Ss* making this type of error on Trial 6 for each of the four lists is shown in Table II. List IV did not involve pairs of stimuli which were similar; so pairs that were in the same relative position in the list as the similar pairs in Lists I, II, and III were used to derive the values for List IV. Twenty-nine of the 80 *Ss* who learned List III made at least one error on Trial 6 involving the

response that was the correct response for the similar stimulus in the list. The numbers for Lists I, II, and IV were 18, 11, and 8 respectively. Analysis of the data in Table 2 yielded a chi-square of 19.94 ($p < .001$), with a considerable proportion of that value coming from the two cells under List III. Although this does not prove that similarity was the only factor operating to produce the differences shown in Table 1, it does provide considerable support for similarity as a major contributing factor. It could, in fact, stand alone as support for the relatively more potent role of primary generalization as compared to secondary generalization in influencing the performance of Ss like those used in the study.

DISCUSSION

The results rather strongly indicate that stimulus similarity does have a relatively greater negative effect on word learning when the dimension of similarity involved is that of structure than when it is meaning or sound. This, in turn, suggests that the Gibson theory may be more adequate for word learning when primary generalization is involved than when secondary generalization is involved.

Gibson, herself, seems to have been more concerned with primary generalization than with secondary generalization. This, as indicated by Underwood (1961), is reflected in her choice of stimulus terms for her own studies (Gibson, 1941, 1942). Confusion, however, was created when the results of these studies were cited as support for the theory as a general theory of verbal learning without proper recognition that words may be similar along dimensions that may involve secondary rather than primary generalization and that a theory adequate in dealing with one may

not be adequate in dealing with the other. This, in fact, may be the case with the Gibson theory; it may be adequate in dealing with primary generalization but inadequate when secondary generalization is involved.

In view of the as yet unsettled question of the effect of ability in paired-associate learning, it is interesting to note the significant main effect for ability and the linear relationship between ability level and learning that is evident in the last column of Table 1.

REFERENCES

- BEECROFT, R. S. Verbal learning and retention as a function of competing associations. *Journal of Experimental Psychology*, 1956, **51**, 216-221.
- FELDMAN, S. M., & UNDERWOOD, G. J. Stimulus recall following paired-associate learning. *Journal of Experimental Psychology*, 1957, **53**, 11-15.
- GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, 1940, **47**, 196-229.
- GIBSON, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *Journal of Experimental Psychology*, 1941, **28**, 93-115.
- GIBSON, E. J. Intra-list generalization as a factor in verbal learning. *Journal of Experimental Psychology*, 1942, **30**, 185-200.
- LINDQUIST, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.
- SCHOER, L. A. Effect of intra-list item similarity on paired-associate learning by learners of high and low verbal ability. *Journal of Educational Psychology*, 1963, **54**, 249-252.
- THORNDIKE, E. L., & LORGE, A. The teacher's word book of 30,000 words. New York: Teachers College, Columbia University, 1944.
- UNDERWOOD, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intra-list similarity. *Journal of Experimental Psychology*, 1953, **45**, 143-149.
- UNDERWOOD, B. J. An evaluation of the Gibson theory of verbal learning. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 197-217.

(Received August 11, 1966)

SECOND-TRY RECALL, RECOGNITION, AND THE MEMORY-MONITORING PROCESS

JOSEPH T. HART

University of California, Irvine

The experiment reported assesses how accurately the memory-monitoring process operates as an indicator of memory storage. Ss were administered a general information questionnaire in 3 versions: a speeded recall test, a 2nd-try recall test, and a recognition test. During the first 2 tests Ss made memory-monitoring judgments about those items that they could not recall. The judgments indicated whether they felt they knew the answer (FK judgments) or did not know the answer ($\overline{\text{FK}}$ judgments). A comparison of the proportions of correct answers, on the memory tests that followed, for the FK and $\overline{\text{FK}}$ items showed a significantly greater number of correct answers for the FK items. The results are discussed (a) in terms of the significance of the memory-monitoring process for a fallible retrieval system and (b) in relation to the concept of availability.

The experiment to be reported here is the seventh in a series of experiments that have been conducted on the memory-monitoring process (Hart, 1965a, 1965b). Memory-monitoring is a term used to describe the intervening cognitive process that begins whenever people try to remember something and fail. When they fail they must then decide whether or not the sought-after item is really in their memories. It is this process, which people use to make judgments about what they have in their heads at times when they cannot get an item out of their heads and into a response, that is called the memory-monitoring (or MEMO) process.

Subjectively, the MEMO process is well-known to everyone in the familiar feeling-of-knowing or tip-of-the-tongue experiences that we have every day. When someone asks the name of an acquaintance, we may try very hard to remember and still fail in our efforts, yet remain convinced that we know the sought-after name. The feeling may be so strong that we persist in our efforts to remember far beyond the limits of practicality.

An important question to ask about these familiar and ubiquitous feeling-

of-knowing experiences is, Are they accurate? This question is important because the MEMO process would be of value only if it could serve as a relatively accurate indicator of the presence or absence of items in the memory storage system. An accurate MEMO process could contribute to the efficiency of a memory system by signaling when continued efforts to retrieve an item would be useful and when they would be useless. In contrast, an inaccurate MEMO process would be detrimental to the overall efficiency of a memory system since such a system would often attempt to retrieve items that were not in storage and not attempt to retrieve items that were in storage.

Recently a method has been developed to assess the accuracy of the MEMO process by applying a simple recall-judgment-recognition paradigm.

This paradigm makes use of the fact that recognition memory usually exceeds recall. Subjects (Ss) are first asked to attempt to recall some memory items; for those items they cannot answer they are required to give a memory-monitoring judgment about whether or not they feel they know

the correct answer well enough to recognize it among several wrong alternatives. After completing the recall test, Ss are given a multiple-choice recognition test covering the same items that appeared on the test of recall.

If the memory monitor is an accurate indicator of memory storage, Ss should do better on those recognition items which they felt they knew but could not recall than on those items they felt they did not know. Accuracy can be easily assessed by comparing the proportion of correct recognitions on FK (feeling-of-knowing) items with the proportion correct on $\bar{FK}$ (feeling-of-not-knowing) items for each S.

The general finding, so far, has been that the memory-monitoring process does serve as a relatively accurate indicator of memory storage. Across seven experiments, the proportion of correct recognitions for the FK items has averaged .61, while the $\bar{FK}$ average has been .41. This result has held for both the recall and recognition of general information questions (such as, "Which planet is the largest in our solar system?") and for the recall and recognition of recently learned paired associates (such as, "WHITE-CXJ").

The experiment reported in this paper extends the investigation of MEMO accuracy, by seeing if Ss can predict their successes and failures when they are given a second chance to recall items that previously eluded them. The recall-judgment-recognition paradigm is altered to insert a second chance to recall and a second chance to make MEMO judgments before the recognition test.

METHOD

Subjects

The eighteen Ss in the experiment were undergraduates at the University of Cali-

fornia, Irvine; all were enrolled in a course on Personality and Cognition.

Materials

Seventy-five general information questions were used as the test materials; these questions had been used in previous experiments (Hart, 1965a, 1965b). "What is the capital city of New Mexico?" is a question from the test. On the recognition test this question appeared with the alternatives: "a. Albuquerque b. Santa Fe c. Los Alamos d. Carlsbad." All of the recognition test questions gave four alternatives.

Procedure

The procedure was divided into three parts. In Part 1 Ss were administered a speeded version of the test of recall. They were given only 3 seconds to answer each question as it was read to them. The instructions read, "Work quickly, provide an answer to every question even if you must guess or give an association, circle the answers you believe might be correct and make feeling-of-knowing judgments (FK or $\bar{FK}$) for the uncircled items." It is important to note that Ss were required to provide some kind of answer to every question, even if they felt that they had no idea about what the correct answer might be. This instruction was necessary because if Ss are allowed to leave answers blank they sometimes withhold answers about which they are not certain and assign an FK judgment to those items. Overcautious withholding of answers can result in an inflation of the FK accuracy score.

In Part 2 Ss followed the same procedures, except that they were given a printed form of the questions and allowed to pace themselves. Also, in Part 2 they were asked to consider, when making their MEMO judgments, whether they would be able to recognize the correct answers when they saw them. For Part 1 the criterion question Ss asked themselves when making MEMO judgments was, "Even though I can't remember the answer at the moment, could I remember if given more time?" For Part 2 the criterion question was, "Even though I can't produce an answer now, do I know it to the extent that I could pick the correct answer from among several wrong answers?"

In Part 3 of the experiment, Ss were given 10 minutes to complete the multiple-choice recognition test on the 75 items. The Ss were instructed to answer all the questions, guessing when necessary.

RESULTS

MEMO accuracy is evaluated by seeing how well Ss can predict which additional items they will recall when given a second try and which additional items they will recognize. To do this it is necessary to show that most Ss do improve their scores on the second-try recall and recognition tests. Otherwise there would be no memory additions for the MEMO process to predict.

In their first attempts to answer the 75-item questionnaire, Ss averaged 24.1 correct answers ($SD = 8.2$). In their second try at recall Ss averaged 27.6 correct answers ($SD = 7.9$), a small but significant improvement. All but one of the Ss improved on the second try. (One S actually missed one more item when he was given a second chance!) On the recognition test, however, even that errant S showed improvement; the average was 46.1 items correct ($SD = 5.9$, with a range of 11–28 additional items correct in moving from the second-try test of recall to the recognition test).

To evaluate MEMO accuracy we examine only those items that the Ss missed and believed that they missed on the preceding memory test. Those are the items for which Ss will have made FK and $\overline{\text{FK}}$ judgments. We then look at how well those judgments predicted the actual memory performances for the memory test that followed.

If the MEMO process is functioning with at least a minimal level of accuracy then many more of the correct additions should have been judged FKs than $\overline{\text{FK}}$ s. Or, in other words, people should eventually remember more items that they feel themselves to know (but cannot recall) than items they feel they don't know. The proportion of $\overline{\text{FK}}$ items

subsequently gotten correct can be compared to the proportion of correct FK items to evaluate MEMO accuracy.

In moving from the first test of recall to the second test, the mean proportion of FK items correct was .25 ($SD = .25$); for $\overline{\text{FK}}$ items the proportion correct was .05 ($SD = .03$). A t test of these means (for correlated scores) yields a value of 3.8, which is significant at the .01 level for 17 df . There were no reversals; for all Ss the FK proportion exceeded the $\overline{\text{FK}}$ proportion.

In moving from the second-try test of recall to the recognition test the mean proportion of FK items correct was .59 ($SD = .14$); for $\overline{\text{FK}}$ items the proportion correct was .30 ($SD = .09$). A t test for these means gives a t value of 6.6, which is significant at the .001 level for 17 df . All the Ss showed more FK recognitions than $\overline{\text{FK}}$ recognitions; the reported means indicate that nearly twice as many FK items were recognized as $\overline{\text{FK}}$ items. The FK proportion, .30, does not differ significantly from the .25 value to be expected if the Ss merely guessed on the four-choice recognition items.

It appears, from these comparisons, that Ss can tell what they have in their memory systems even before they can produce the memories. Clearly, however, the MEMO process is not perfectly accurate. If it were, the $\overline{\text{FK}}$ proportion would be 1.00 and the FK proportion, on the second test of recall, would be .00. Instead Ss were able to recall only .25 of the items that they believed they might produce if given a second chance. What happened was that many Ss overestimated the number of items they would be able to remember if given a second chance.

If we look at the MEMO judgments made on the first-try recall test and use them to predict recognition responses rather than second-try recall then the FK mean rises to .62 ($SD = .18$) and the $\overline{FK}$ mean to .38 ($SD = .08$). A t test for correlated scores yields a value of 4.85, which is significant at the .001 level for 17 df . The monitor appears, from these results, to be a more accurate predictor of recognition performances than recall performances.

Since the directions to the Ss did not specify how much time they would be given on their second attempts it is possible that they were misled. Given more time, perhaps they would be able to retrieve more of the FK items. This possibility will be tested in further experiments.

DISCUSSION

The results show that the MEMO process can be used as a relatively accurate indicator of memory storage. The indicator can be applied to the prediction of recognition successes and failures, and, somewhat less accurately, to the prediction of second attempts at recall. When people feel that they know something, it is very likely that they do know it, and when they feel that they do not, it is likely that they do not.

An appreciation of the possible significance and usefulness of the MEMO process can be gained by comparing the human information-processing system with that of computers. Human systems are enormously flexible but they lack speed and, most importantly, they are fallible. It is this fact of memory fallibility that makes the MEMO process important. If human beings were computers with nearly infallible memories they would not need a MEMO indicator. A memory item, if in storage,

would be retrieved; failure to retrieve an item would simply mean that it was not in storage.

Clearly, however people do not function this way—retention often exceeds recall. For a fallible memory system, in which what is retained often exceeds what is retrieved, the MEMO process can be exceedingly useful. It can operate as a storage indicator at times when retrieval efforts are unsuccessful. When the indicator signals that an item is not in storage, the system need not continue to expend useless effort and time at retrieval. Instead, input can be sought that will put the item into storage. Or, if the MEMO indicator signals that an item is in storage, the system can persist in attempts to retrieve information, avoiding the redundant input of information that is already possessed.

Apart from its practical usefulness to the human memory system, the MEMO process is also of interest because of the theoretical questions it raises about the system. The demonstration that people can rely upon subjective experiences to predict the contents of their memories is of general interest to psychologists concerned about the functional relationships between subjective processes and behavior processes.

For example, recent studies of verbal learning have emphasized the distinction between response learning and association learning (Underwood & Schultz, 1960). One interesting theoretical construct that has developed from this distinction is the concept of availability. This concept asserts that associations may be present but production of the associations can be impeded until the responses are made available.

The concept has been experimentally applied to show that the usual asymmetry between forward and

backward associations is removed if the backward associate is made as available as the forward. When unavailable responses are made available hidden associations become visible in overt behavior (Horowitz, Norman & Day, 1966).

These experiments on availability indicate that, "Even if a response is not available—not overt—it is activity in the nervous system [Horowitz, 1965, p. 8]." The point here is that a psychology of verbal learning must certainly examine carefully any subjective process which allows people to know that an association is present at times when the response is unavailable. The memory-monitoring process does just that.

REFERENCES

- HART, J. T. Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 1965, **56**, 208-216.
- HART, J. T. *Recall, recognition, and the memory-monitoring process*. Unpublished doctoral dissertation, Stanford University, 1965.
- HOROWITZ, L. M. Latent verbal learning. Paper read at Symposium on Verbal Learning, American Association for the Advancement of Science, Berkeley, December 1965.
- HOROWITZ, L. M., NORMAN, S. A. & DAY, R. S. Availability and associative symmetry. *Psychological Review*, 1966, **73**, 1-15.
- UNDERWOOD, B. J., & SCHULTZ, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.

(Received July 22, 1966)

THE MANY FACES OF INTELLIGENCE¹

CHARLES E. WERTS

National Merit Scholarship Corporation

Data on a sample of 127,125 college freshmen were used to study the relationship between high school grade average and various types of extracurricular talent displayed during high school. In the scientific, literary, leadership, speech, drama, music, and art areas, the percentage of achievers was greater among students with high grades than among students with low grades in high school. Students with high grades usually won recognition in several of these extracurricular areas, whereas the majority of students with low grades did not show any extracurricular achievement.

High school grades and aptitude-achievement test scores are the most widely used criteria for college admission. In recent years, a variety of critics have argued that reliance on these criteria results in heavy loss of persons who have the capability of "creative" performance in "real life" (nonacademic) situations. These critics (typified by Holland and Richards, 1965) point to the low correlations of academic achievement tests with nonacademic accomplishments as evidence that selection by academic performance necessarily will miss a large proportion of persons who have the capability of "creative" or "real life" performance. The validity of this argument will be examined here for those cases in which correlations are computed on samples of college students.

METHOD

Subjects

The subjects were 127,125 students entering 248 4-year colleges and universities in

the fall of 1961. With few exceptions, these students included the entire freshman class at each institution. The colleges were chosen to include a wide variety of types of institutions in all regions of the United States. Details of the college selection procedure were given by Astin (1965).

Measures

Along with the usual registration forms, each student filled out a short information form which included the following questions:

1. Circle one: Male Female
2. Your high school average (circle one):

D	C	C+	B-	B	B+	A-	A	A+
1	2	3	4	5	6	7	8	9

3.

Indicate whether you have achieved any of the following by underlining the appropriate words. On the line before any item you underline, indicate the number of times you have achieved it.

First, second, or third place in: school science contest; regional or state science contest; national science contest

..... leads in high school or church sponsored plays; first, second, or third in regional or state speech or debate contest; first, second, or third in national speech or debate contest

..... elected to one or more student offices; elected president of my class; received award or special recognition for leadership of any kind

..... participated in national music contest; received a rating of "good" or "excellent" in: state music contest; national music contest

..... won a prize or award in art compe-

¹ This study is part of the National Merit Scholarship Corporation research program which is supported by grants from the National Science Foundation, the Ford Foundation, and the Carnegie Corporation of New York. This paper is an extension of an earlier discussion by Astin (1962) using the same data.

tition (sculpture, ceramics, painting, etc.); exhibited or performed a work of art (painting, musical composition, sculpture) at: my school; place other than my school

.....edited school paper or literary magazine; had poems, short stories, or articles published in public newspaper or magazine (not school paper) or in state or national high school anthology; won literary award or prize for creative writing

Davidson (1963) found that self-reported grades are reasonably accurate and correlate .92 with school-reported grades. Data on a subsample of 20,000 students who participated in the annual National Merit talent search indicate that high school grade average as reported by the students in this sample is linearly related to National Merit Scholarship Qualifying Test (Science Research Associates, 1966) scores, the correlation among college freshmen being about .50.

Design

Each person's responses to the 18 talent items were categorized into a some versus none dichotomy. The number of persons with a given high school grade average (HSG) that checked each item was computed and divided by the total number of persons with that grade average to obtain the percentage of achievers for each HSG category. Analyses were done separately for males and females because their talent patterns are dissimilar, and because among entering freshmen, girls have higher grades ($M = 5.5$) than boys ($M = 4.7$).

Methodological Considerations

The percentages shown in Tables 1 and 2 must be interpreted in light of the following considerations:

1. Because the sample was composed of entering college freshmen, McNemar's (1964) criticism of other studies for restriction in range on ability and criteria applies to this study too. For instance, the validity of comparing A+ students, most of whom go to college, with C students depends mainly on how close the percentage of achievers (nonacademic) among C students who go to college is to the percentage of achievers among representative C students in high school. C students (or perhaps any students) who have these extracurricular achievements may be more likely to go to

college than those who do not. If so, the results will err on the conservative side. The percentages for drama, art, and music may be even more questionable, because a large number of the most talented students in these areas go on to specialized schools—like Juilliard School of Music—rather than to 4-year colleges. A sample of college students really should not be used to study talent loss resulting from use of academic criteria for college admission, since everyone in the sample has survived the selection process. Talent loss from this source might better be studied by comparing those who were selected against those who were not.

2. The use of high school grade average as an academic performance measure is crude, because grades reflect nonintellectual factors like the student's relationship with the teacher; and grades from different high schools and courses frequently are non-comparable due to variations in quality of the student body and in grading practices. Since these factors probably are sources of unreliability, the results again are likely to be conservative. A better design would include both grades and test scores as academic measures.

3. Even if the above limitations were not present in this study, it would be difficult to interpret changes in percentages of achievers with changes in HSG level as meaning there is or is not a necessary connection between these achievements and HSG. The problem can be illustrated by the finding that 17.6% of the A+ males, compared with 1.1% of the C males, checked the regional science item. What happened to the 82.4% of the A+ and the 98.9% of the C students who did not win? Perhaps all A+ but only a small portion of C students who participated won prizes—or perhaps all C students won (assuming only this 1.1% of C students participated), and only 17.6% of A+ students (assuming all A+ students participated). To assume that the ratio of 17.6% to 1.1% represents the true association between HSG and winning the regional science contest is to assume automatically that the percentage of all A+ students competing is the same as the percentage of all C students who competed. If almost all A+ students and only a small percentage of C students competing in the regional science contest won, the ratio of 17.6% to 1.1% might be a radical underestimate of the true association between HSG and winning the contest. Similarly, if the reverse were true, the 17.6% to 1.1% ratio might be a gross overestimate of the

true association. Even knowledge of which persons were formally known to be entering the contest would not solve the problem of interpretation, since it would not be known how HSG affects which persons formally enter the contest. Thus, almost regardless of the percentage differences between HSG groups, little can be said about the true association between HSG and winning regional science contests. A more appropriate manner of studying how intelligence (or HSG) is related to talented performance has been suggested by McNemar (1964). In his design, IQ scores are plotted against a criterion measure, such as literary creativity. The crucial point is that McNemar conceives of creativity as a variate ranging from high to low that should be measured for persons in a certain line of endeavor: In other words, writing creativity should be measured for writers, scientific creativity for scientists, etc. Thus, studying the percentage of state science winners at various HSG levels errs in comparing state science winners with an inappropriate reference group—students in general.

A related problem lies in interpreting the association between a specific accomplishment and academic performance measures—even if the data met the assumptions required for the appropriate coefficient, and the coefficient of association had been computed on an unbiased sample with comparable academic performance measures and accurate talent data. The following example illustrates one difficulty. Suppose there were 100 high-HSG persons each of whom displayed different talents, and that this group were compared to 100 low-HSG, talentless persons. Since only high-HSG persons would have shown any talent at all, one probably would conclude that HSG is closely related to talented performance. However, if a phi coefficient were computed on the whole sample between a specific accomplishment (yes or no) and HSG (high versus low) a low coefficient of .07 would result, since for each accomplishment 99 of the high-HSG persons would not have displayed that accomplishment. Thus, a list of 100 small phi coefficients, one for each accomplishment, could be obtained. Because talent correlates perfectly with HSG in this example, the small size of each phi coefficient is only a reflection of the small proportion of high-HSG persons who had that particular talent. The size of the phi coefficient tells nothing about the true association between HSG and that particular talent. The important point is that in

actuality the correlation of HSG with a given accomplishment will be influenced greatly by the proportion of the whole population that engages in such a line of endeavor. It follows that a low correlation between HSG and a given accomplishment (computed on a broad sample) cannot be validly interpreted to mean that HSG is not related to that particular accomplishment, nor can a list of such low correlations for different talents be offered as evidence that HSG is not, in general, associated with talented performance. The same interpretive problem exists for studies that correlate (on a broad sample) aptitude or intelligence test scores with various types of accomplishments. One can substitute the word "aptitude" or "intelligence" for "academic performance" in this methodological discussion.

The above arguments indicate that the correlations of specific achievements with HSG are not meaningful in themselves. The example in which 100 low phi coefficients were obtained suggests one interpretation: Given the list of phi coefficients, one can deduce that talent in general and HSG are perfectly correlated by using a multiple correlation model. Theoretically, the multiple correlation model implies that the skills inherent in high HSG can be utilized in a variety of extracurricular achievements. The correlation of one type of achievement with HSG cannot in itself be interpreted, precisely because some persons will display their academic skills in another type of achievement.

RESULTS

The percentage of achievers (persons checking an item) at each HSG level is shown in Table 1 (males) and Table 2 (females). The percentage of achievers rose exponentially with rise in HSG for the three science and the three literary items. These percentages plotted as a straight line on semilog paper. Curves for other items showed a generally rising percentage of achievement with rise in HSG.

A different perspective was gained by looking at the number of areas in which a person achieved. Using the item grouping indicated in Table 1, there were seven possible areas of achievement: Science, Drama, Speech

TABLE 1
PERCENTAGE OF MALES CHECKING EACH ACHIEVEMENT AT VARIOUS HIGH SCHOOL GRADE LEVELS

Achievement area and item no.	Percentage checking item (base rate) ^a	High school grade average									Correlations	
		D	C	C+	B-	B	B+	A-	A	A+	r_{bis}^b	r_{HR}^b
Science												
1. School science award	9.4	3.1	4.2	5.7	7.1	8.5	12.1	15.6	20.1	31.6	.16	
2. Regional science award	3.5	.8	1.1	1.9	2.7	3.1	4.2	6.5	8.6	17.6	.12	.15
3. National science award	.3	.0	.1	.1	.2	.2	.4	.5	.7	1.5	.04	
Drama												
4. Lead in school play	22.3	10.2	14.4	17.6	21.3	23.3	26.2	29.1	31.8	32.4	.12	
Speech and debate												
5. Regional speech award	4.6	1.2	2.0	2.7	3.5	4.1	5.6	8.4	10.1	11.2	.11	.08
6. National speech award	.2	.3	.2	.1	.2	.2	.4	.3	.4	.6	.02	
Leadership												
7. Elected to student office	37.6	14.9	23.2	28.1	33.4	39.1	45.0	53.0	58.2	60.9	.22	
8. Elected class president	13.1	3.9	6.5	8.4	11.0	13.4	17.2	20.6	22.0	20.5	.15	.21
9. Leadership award	27.5	11.0	14.2	17.6	22.9	27.0	35.6	42.8	49.6	56.8	.24	
Music												
10. National music participant	1.5	.8	1.3	1.5	1.7	1.5	1.5	1.6	1.6	2.8	.01	
11. State music award	8.3	4.1	5.6	6.8	7.8	8.1	9.5	11.4	12.1	11.0	.07	.02
12. National music award	.6	.7	.5	.4	.5	.6	.7	.7	.9	1.5	.02	
Art												
13. Art prize	3.4	3.8	3.0	3.1	3.5	3.5	3.7	3.1	3.5	6.6	.01	
14. School art exhibit	4.4	3.6	3.9	3.8	4.5	4.6	4.3	4.7	5.2	7.5	.06	-.05
15. Other art exhibit	3.6	2.6	2.8	3.0	3.6	3.6	3.9	4.5	4.6	6.1	.03	
Literary												
16. Editor of school publication	8.2	2.0	3.5	4.6	6.5	7.5	10.6	13.8	18.5	21.4	.15	
17. Published other than school	6.4	2.0	3.1	4.0	4.9	5.9	8.4	10.7	13.2	18.9	.12	.10
18. Creative writing award	3.2	.8	1.0	1.4	2.0	2.8	4.0	6.3	8.5	12.7	.12	
Number of areas checked												
None		64.5	51.4	44.1	36.5	31.4	23.0	16.2	11.3	7.6		
More than one		12.8	18.0	23.4	29.0	34.0	42.7	50.8	60.3	70.4		

Note.—For each achievement item, the percentage of persons checking the item at each of the nine high school grade levels is shown. The actual number of males at each high school grade level, going from left to right, is: 609, 9,479, 12,711, 12,192, 14,468, 11,988, 7,377, 4,830, and 655.

^a The base rates were calculated on the 76,015 males, of whom 1,706 did not respond to the HSG question. For this reason the base rate is not a weighted average of the percentages for the nine HSG levels.

^b The point-biserial correlations (r_{bis}) of each achievement item with high school grade average were computed in order to make comparisons with equivalent correlations from Holland and Richards' (1965) Table 2. It can be seen that the point-biserial correlations within each of the six achievement areas are close to the correlations (r_{HR}) reported by Holland and Richards for that area.

and Debate, Leadership, Music, Art, and Literary. The percentage of persons at each HSG level that checked none of the achievement items may be compared with the percentage that checked items in more than one area of achievement. The percentage checking no item decreased systematically from 64.5% at the D level to 7.6% at the A+ level for males and from 47.2% to 5.6% for females, whereas only 12.8% of the D versus 70.4% of the A+ males (15.7% versus 72.6% for females) achieved in more than one area. If these nonacademic

achievements were to be used for selection purposes, the major effect would likely be the elimination of the majority of the low grade students.²

The achievements items used here were similar to those used by Holland and Richards (1965), in which point-biserial correlations were given for

² In a personal communication (June 6, 1966) McNemar suggested that these data indicate "that degree of versatility in accomplishments is a function of intellectual level."

TABLE 2
PERCENTAGE OF FEMALES CHECKING EACH ACHIEVEMENT AT VARIOUS HIGH SCHOOL GRADE LEVELS

Achievement area and item no.	Percentage checking item (base rate) ^a	High school grade average									Correlations	
		D	C	C+	B-	B	B+	A-	A	A+	r_{bis}^b	r_{HR}^b
Science												
1. School Science Award	6.9	1.1	3.2	3.6	4.5	5.5	7.0	9.2	12.4	14.6	.12	
2. Regional Science Award	2.7	.0	.8	1.2	1.6	1.9	2.7	3.9	5.3	6.7	.09	.10
3. National Science Award	.2	.0	.1	.1	.1	.1	.2	.1	.5	.8	.02	
Drama												
4. Lead in School Play Speech and debate	27.3	7.9	17.4	20.8	23.8	25.0	29.5	31.4	34.6	32.7	.11	
5. Regional Speech Award	5.8	2.3	2.6	3.6	4.0	4.6	6.3	7.8	9.2	11.3	.09	.10
6. National Speech Award	.2	.0	.1	.1	.1	.3	.2	.2	.3	.8	.02	
Leadership												
7. Elected to Student Office	48.9	16.9	25.9	31.5	40.1	44.9	53.0	59.8	64.7	72.4	.25	
8. Elected Class President	6.3	3.4	3.2	3.6	4.4	5.2	6.5	8.5	10.2	10.4	.09	.20
9. Leadership Award	35.7	11.2	15.0	17.4	23.4	28.7	39.9	49.8	55.6	62.6	.30	
Music												
10. National Music Participant	2.9	1.1	2.9	2.8	2.9	3.0	2.8	3.0	3.0	3.1	.00	
11. State Music Award	12.8	11.2	7.9	9.8	10.2	12.0	13.8	15.6	15.7	14.7	.07	.06
12. National Music Award	1.3	.0	1.0	1.1	1.0	1.1	1.3	1.4	1.7	1.9	.02	
Art												
13. Art Prize	7.0	7.9	5.2	6.6	7.5	7.1	7.1	7.1	7.3	9.6	.01	
14. School Art Exhibit	9.6	10.1	8.2	8.8	10.0	9.7	9.7	9.7	9.8	10.5	.01	-.05
15. Other Art Exhibit	8.2	4.5	5.7	6.8	7.8	8.3	8.4	9.1	9.2	9.9	.04	
Literary												
16. Editor of School Publication	16.4	3.4	5.4	7.5	10.2	12.6	17.9	22.9	27.7	31.8	.02	
17. Published Other Than School	11.3	3.4	4.5	6.0	8.1	8.9	12.0	15.3	18.3	26.4	.15	.16
18. Creative Writing Award	6.7	2.3	1.8	2.3	3.8	4.3	6.8	10.3	13.1	18.3	.16	
Number of areas checked												
None		47.2	44.1	36.7	29.1	24.7	16.8	11.6	8.4	5.6		
More than one		15.7	23.5	30.0	37.4	42.4	52.4	60.8	66.8	72.6		

Note.—For each achievement item, the percentage of persons checking the item at each of the nine high school grade levels is shown. The actual number of females at each high school grade level, going from left to right, is: 89, 2,270, 4,254, 8,959, 11,292, 11,459, 8,266, 5,678, and 645.

^a The base rates were calculated on 51,110 females, of whom 1,198 did not respond to the HSG question. For this reason the base rate is not exactly the weighted average of the percentages for the nine HSG levels.

^b The point-biserial correlations (r_{bis}) of each achievement item with high school grade average were computed in order to make comparisons with equivalent correlations from Holland and Richards' (1965) Table 2. It can be seen that the point-biserial correlations within each of the six achievement areas are close to the correlations (r_{HR}) reported by Holland and Richards for that area.

each of six areas of achievement: Science, Leadership, Dramatic Arts, Artistic, Literary, and Musical. Items 1, 2, and 3 in the present study fell into Holland and Richards' Science area; Items 4, 5, and 6 into their Dramatic Arts area; Items 7, 8, and 9 into Leadership; Items 10, 11, and 12 into Musical; 13, 14, and 15 into Artistic; and 16, 17, and 18 into Literary. Tables 1 and 2 show that for each area Holland and Richards obtained correlations of HSG with achievement that came close to the

three point-biserial correlations for corresponding items in the present study. This finding is not surprising, since both studies used survey data on large samples of college students.

The point-biserial correlations for the different items are in part non-comparable, because each correlation is not free to vary within the same range. The maximum obtainable correlation for each achievement item was computed from the original cross tabulations by maintaining the marginal row and column frequencies and

arbitrarily assigning to the achievers the highest possible grades, so that no nonachiever was assigned a higher HSG level than an achiever. Thus, it was possible to calculate the maximum point-biserial correlations for Table 1, going from Item 1 to Item 18: .57, .38, .12, .73, .42, .11, .81, .63, .77, .27, .55, .18, .37, .41, .38, .55, .50, and .36. Each point-biserial correlation was divided by its maximum value, as is often done for the phi coefficient. The point-biserial correlations for the scientific, literary, and leadership items were about 25% to 35% of their maximum, whereas the artistic and musical item correlations averaged less than 10%, with dramatic arts items somewhere in between. Although it is tempting to suggest that the ratio of the biserial correlation to its maximum is a fairly good measure of the degree to which academic skills are important for a particular achievement, this interpretation ignores the problems raised earlier.

A multiple correlation of .38 for males and .37 for females was obtained between the 18 achievement items and HSG. The multiple correlation model simulates the "real life" situation in which the skills intrinsic to high HSG can be utilized in literary and/or scientific and/or leadership and/or other types of achievements. However, the data did not fit the assumptions for multiple correlation in at least two ways: (a) Some of the regression lines were nonlinear, and (b) according to McNemar, the underlying distribution for a given type of achievement plotted against HSG probably is skewed. Perhaps more serious were the methodological problems that arose as a result of inadequate coverage of types and degrees of achievements. In the same sense that the range in verbal ability is not adequately measured by use of

only very difficult test items, various degrees of talent must be included or the size of the multiple correlation necessarily will be limited. For instance, the coverage of science achievements in the present data appeared inadequate, since even the most frequently checked item, winning a high school science award, probably did not tap large numbers of students with considerable scientific ability. Furthermore, many other varieties of scientific achievement, such as publishing a scientific article, building an intricate device, and inventions, were not covered. Considering that 18 items were probably an inadequate sample of nonacademic achievements, it seems reasonable to suggest that with more adequate coverage it might have been possible to predict HSG almost as well from nonacademic achievements as from National Merit Scholarship Test Scores—which correlated about .50 with HSG. The multiple correlation of HSG with nonacademic achievements may tell something about the overlap of skills between academic and nonacademic achievements; however, it tells little about talent loss in general, because different types of talent vary widely with respect to the kinds of skills necessary for their successful demonstration.

A low correlation between a particular nonacademic accomplishment and academic achievement cannot be validly interpreted to mean that academic ability is not necessary for that accomplishment. Therefore, one should not cite these low correlations as evidence that use of academic performance criteria—grades and achievement test scores—for college admission results in heavy loss of persons who are capable of "creative" performance in "real life" (nonacademic) situations.

REFERENCES

- ASTIN, A. W. A note on the relationship between creative and academic achievement. Paper read at American Association of College Registrars and Admission Officers, Kansas City, April 1962.
- ASTIN, A. W. *Who goes where to college?* Chicago: Science Research Associates, 1965.
- DAVIDSEN, O. M. Reliability of self-reported high school grades. Unpublished research report, American College Testing Program, 1963.
- HOLLAND, J. L., & RICHARDS, J. M., Jr. Academic and nonacademic accomplishment: Correlated or uncorrelated? *Journal of Educational Psychology*, 1965, **56**, 165-174.
- MCNEMAR, Q. Lost: Our intelligence? Why? *American Psychologist*, 1964, **19**, 871-883.
- SCIENCE RESEARCH ASSOCIATES. *National merit scholarship qualifying test, Spring 1966, interpretive manual*. Chicago: Author, 1966.

(July 25, 1966)

THE MANY FACES OF TALENT: A REPLY TO WERTS

JOHN L. HOLLAND AND JAMES M. RICHARDS, JR.
American College Testing Program

Werts (1967) (a) argues that our correlational studies are misleading because of various statistical and sampling artifacts, and (b) implies that a heavy reliance on high school grades in selection would result in no great loss of students with other kinds of talent. When his data are retabulated, they reveal that grades are an inefficient way to select for nonacademic talents and have little relationship with other talents. His explanations for the negligible correlations between academic and nonacademic accomplishment are contradicted by the published evidence.

In a series of reports, we have found little or no relationship between academic and nonacademic accomplishment. In addition, we found that measures of academic potential are good predictors of academic success and measures of nonacademic potential are good predictors of nonacademic accomplishment—but not vice versa (Holland & Nichols, 1964; Holland & Richards, 1965; Richards, Holland, & Lutz, 1966a, 1966b). From these results, and similar findings by others (Locke, 1963; Wallach & Kogan, 1965), we have concluded that if sponsors—college admissions offices, scholarship agencies, etc.—rely heavily on measures of academic potential, they will miss many people with nonacademic talents.

In the preceding report, Werts (1967) (a) argues that our correlational studies are misleading because of various statistical and sampling artifacts, and (b) implies that a heavy reliance on high school grades in selection would result in no great

loss of students with other kinds of talent. We will try to show that Werts's own data do not support his hypotheses, and that the published evidence contradicts his hypotheses.

WERTS'S DATA REANALYZED

In his Tables 1 and 2, Werts tries to demonstrate that use of high school grades in selection will produce a group of students with many talents. For example, in Table 1 the selection of only A+ students would result in a group of men students of whom 17.6% have won regional science awards, whereas only 3.5% of men students in general have won such an award.

But the question of talent loss involves what you *miss* by various selection rules as well as what you *get*, so that Werts's table is only partially relevant to the question at issue. Werts's results do indicate clearly that even low correlations yield impressive success ratios if the selection ratios are stringent, but this occurrence is neither new nor surprising.

By reanalyzing Werts's data, however, it is possible to determine the talent loss resulting from exclusive use of grades in selection.¹ From his data, we created a single table that shows what percentages of students with

¹The percentages reported in Table 1 were computed by multiplying the percentage of students at each grade level with a given kind of achievement by the number at that grade level and summing across grade levels to obtain the total number of students with that kind of achievement.

TABLE 1
PERCENTAGES OF TALENTED STUDENTS ELIMINATED BY THE USE OF DIFFERENT
GRADE LEVELS IN SELECTION

Achievement area	Males				Females				All students			
	B+	A-	A	A+	B+	A-	A	A+	B+	A-	A	A+
Science												
School science award	46.1	66.8	83.2	97.1	31.8	54.7	76.4	97.3	41.4	62.8	81.0	97.1
Regional science award	42.7	61.8	79.9	95.6	28.0	50.8	74.6	96.8	37.7	58.1	78.1	96.0
National science award	36.8	60.3	78.4	95.1	26.4	52.9	62.1	94.3	33.7	58.1	73.5	94.8
Drama												
Lead in school play	57.8	76.6	89.5	98.7	40.4	65.1	84.1	98.5	50.0	71.4	87.1	98.6
Speech and debate												
Regional speech award	45.7	65.4	83.6	97.9	33.1	57.7	79.7	97.5	39.9	61.8	81.8	97.7
National speech award	48.3	75.0	87.2	97.8	42.6	63.9	79.6	95.4	46.2	70.8	84.4	96.9
Leadership												
Elected to student office	55.5	74.7	88.6	98.6	38.3	63.0	83.1	98.1	47.5	69.2	86.0	98.4
Elected class president	51.1	72.2	87.7	98.6	34.0	57.5	79.6	97.9	46.9	68.6	85.8	98.4
Leadership award	50.4	71.2	86.5	98.2	31.9	57.3	80.2	97.8	41.7	64.7	83.6	98.0
Music												
National music participant	65.4	81.3	91.6	98.4	47.9	69.9	87.0	98.6	55.6	74.9	89.0	98.5
State music award	57.3	75.8	89.4	98.8	40.0	64.6	84.7	98.5	48.5	70.1	87.0	98.7
National music award	56.9	76.1	87.9	97.7	40.4	64.2	82.6	98.1	47.2	69.1	84.8	97.9
Art												
Art prize	64.9	82.5	91.6	98.3	46.9	69.9	86.5	98.2	54.4	75.1	88.6	98.3
School art exhibit	64.3	80.1	90.8	98.5	47.1	70.3	87.0	98.6	54.0	74.2	88.5	98.5
Other art exhibit	60.4	77.9	90.2	98.5	44.2	67.6	85.8	98.4	50.6	71.6	87.5	98.5
Literary												
Editor of school publication	45.8	66.5	83.1	97.7	30.2	55.2	78.3	97.5	36.9	60.1	80.4	97.6
Published other than school	47.0	67.8	84.2	97.4	32.4	56.6	78.8	97.0	39.1	61.7	81.3	97.2
Creative writing award	39.2	59.5	79.1	96.5	25.5	48.8	74.2	96.5	31.2	53.2	76.3	96.5
Number of areas checked												
Greater than one	51.9	72.0	86.7	98.2	36.6	61.5	82.3	98.1	44.5	66.9	84.6	98.1
Academic achievement												
Top decile	0.0	0.0	26.2	91.2	0.0	0.0	0.0	87.0	0.0	0.0	4.9	89.5
All students	66.6	82.7	92.6	99.1	47.8	70.8	87.3	98.7	59.0	77.9	90.5	99.0

Note.—This table is based on a reanalysis of data presented by Werts (1967). Entries are the percentage of students with each kind of achievement who have grades lower than the column head.

various kinds of achievement are eliminated by the use of various grade levels as selection scores. We have not considered grade levels below B+ because below that level a college usually has a nonselective admissions policy. When you look at Table 1, it is clear that the selection of only A+ or A students (a selection rule that will admit nearly all students in the top decile on grades) will result in the elimination of 74-93% of all students with various kinds of nonacademic accomplishment. To take another more concrete example, if you only select the A or A+ students (about the top decile of *academic* talent), you would get 1,843 class presidents, but you would miss 11,096 class presidents. If you select only students with B+ grades or higher you reject 51.9% of the men and 36.6% of the women who have checked

more than one kind of achievement. In short, the use of grades as an efficient sign for the selection of multi-talented persons is not warranted by the Werts data.

MORE EVIDENCE

In this section, we will show that the published evidence strengthens the hypothesis that academic and non-academic accomplishment are largely independent dimensions, and that the numerous replications that support this hypothesis cannot be invalidated by the peculiarities of the data or the statistical problems that Werts (1967) and McNemar (1964) have suggested.

1. Do the small percentages of students with nonacademic accomplishments present a misleading picture of the actual relationships between academic and nonacademic accomplishment? Our Table 1 indicates that this

argument has a *little* validity, since in only a few cases is the percentage of nonacademic achievers eliminated quite as large as the percentage of all students eliminated. Nevertheless, almost as many nonacademic achievers as compared to all students are eliminated, and many more nonacademic than academic achievers are eliminated.

The most persuasive refutation of the percentage argument is provided in other research (Richards, Holland, & Lutz, 1966a, 1966b) by a special criterion scale developed to assess Recognition for Academic Accomplishment (RAA) at the college level. This scale of only five items, such as being on the Dean's honor list, selected for an honors program, elected to Phi Beta Kappa etc., has more statistical defects than the typical nonacademic achievement scales that Werts criticizes. Its estimated reliabilities range only from .31 to .50; its means and *SDs* for various samples are less than 1 so that it has a violently skewed distribution of student scores. Therefore, if the negligible correlations that are usually found between academic and nonacademic accomplishment are simply due to statistical defects, the RAA scale should also be uncorrelated with grades. Yet the RAA scale is moderately correlated with grades and academic test scores, for both concurrent and predictive relationships, while the nonacademic achievement scales are not correlated with grades, so that the evidence provides both convergent and discriminant validity. In other words, the correlations do present an accurate picture of the actual situation, and the content of these scales is a more important determinant of correlation than are some of their statistical properties. Finally, in the Holland and Richards

(1965) report there are several criterion scales that have normal distributions (Total Competencies and Originality), but they too have negligible relationships with grades.

2. What happens when you use academic and nonacademic predictors? If Werts's hypothesis has validity, then the selection of students by reliance on high school grades would lead to the selection of both academic and nonacademic achievers. Several studies at the National Merit Scholarship Corporation, however, contradict this hypothesis. In a simulated selection study, Nichols & Holland (1964) found that selection by high grades produced a student sample that obtained high grades in college but a poor record of nonacademic accomplishment. In contrast, selection by nonacademic achievements produced a student sample that obtained an outstanding record of nonacademic accomplishment in college without lowering the level of academic performance. Other studies at National Merit by Holland (1961, 1964), Holland and Nichols (1964), Astin (1962), Nichols (1966), and Roberts (1965), without exception, suggest that you will get different groups of students with academic and nonacademic criteria—so different that they are not useful substitutes for one another.

More recently, using large diverse student samples, Richards, Holland and Lutz (1966b) at ACT found that high school grades would not predict nonacademic accomplishment in college, although nonacademic predictors, once again, had useful predictive validities across 22 colleges.

3. Is the lack of relationship between academic and nonacademic accomplishment due to a narrow range of talent? In a representative sample of 18,378 students taken from a population of 612,000 college applicants,

Holland and Richards (1966) again found that the relationships between ACT test scores and high school grades are generally negligible. This refutes the argument that our results are due largely to the restriction in range resulting from the college selection process. In an older article, using a very narrow range of talent (National Merit Finalists), Holland (1961) found that the NMSQT and the SAT correlate as much as .64 and .57 over a period of 1 year, but neither test had any appreciable *concurrent* relationship with scientific or artistic accomplishment outside the classroom (r 's ranged only from $-.19$ to $+.13$). In short, there is no longer any evidence that range of talent is a plausible explanation of all of the repeated occurrence of negligible relationships.

4. What about...? We have done most of the things Werts cites as desirable in studies of academic and nonacademic achievement. In the specific study Werts criticizes, we used both test scores and grades as measures of academic accomplishment, and found that grades produce relationships that are equivalent to those obtained by the use of a more reliable test (see Table 4, Holland & Richards, 1965). We have studied academic and nonacademic accomplishment in a large sample of college applicants (Holland & Richards, 1966) and found again that academic and nonacademic accomplishment have only negligible relationships. Last, we studied each nonacademic accomplishment for a subgroup of students presumably interested in that accomplishment (Holland & Richards, 1966); for example, we studied the relationship between scientific accomplishment and academic accomplishment using only students intending to major in scientific fields. In a few cases, interest does act as a modera-

tor of the relationship between academic and nonacademic accomplishment so that the relationship is noticeably greater within subgroups than for the total group. Even in these cases, however, academic accomplishment is still an inefficient predictor of nonacademic accomplishment.

REFERENCES

- ASTIN, A. W. A note on the relationship between creative and academic achievement. Paper read at American Association of College Registrars and Admission Officers, Kansas City, April 1962.
- HOLLAND, J. L. Creative and academic performance among talented adolescents. *Journal of Educational Psychology*, 1961, **52**, 136-147.
- HOLLAND, J. L. The selection of students for special scholarships. *Journal of Higher Education*, 1964, **35**, 32-37.
- HOLLAND, J. L., & NICHOLS, R. C. Prediction of academic and extracurricular achievement in college. *Journal of Educational Psychology*, 1964, **55**, 55-65.
- HOLLAND, J. L., & RICHARDS, J. M., JR. Academic and nonacademic accomplishment: Correlated or uncorrelated? *Journal of Educational Psychology*, 1965, **56**, 165-174.
- HOLLAND, J. L., & RICHARDS, J. M., JR. Academic and non-academic accomplishment in a representative sample taken from a population of 612,000. ACT Research Report No. 12. Iowa City, Iowa: American College Testing Program, 1966.
- LOCKE, E. A. The development of criteria of student achievement. *Educational and Psychological Measurement*, 1963, **23**, 299-307.
- MCNEMAR, Q. Lost: Our intelligence? Why? *American Psychologist*, 1964, **19**, 871-882.
- NICHOLS, R. C. Non-intellective predictors of achievement in college. *Educational and Psychological Measurement*, 1966, **26**, 899-915.
- NICHOLS, R. C., & HOLLAND, J. L. The selection of high aptitude high school graduates for maximum achievement in college. *Personnel and Guidance Journal*, 1964, **43**, 33-40.
- RICHARDS, J. M., JR., HOLLAND, J. L., & LUTZ, S. W. The assessment of student accomplishment in college. ACT Research Report No. 11. Iowa City, Iowa: American College Testing Program, 1966. (a)
- RICHARDS, J. M., JR., HOLLAND, J. L., &

- LUTZ, S. W. The prediction of student accomplishment in college. ACT Research Report No. 13. Iowa City, Iowa: American College Testing Program, 1966. (b)
- ROBERTS, R. J. Prediction of college performance of superior students. NMSC Research Report No. 5. Evanston, Illinois: National Merit Scholarship Corporation, 1965.
- WALLACH, M. A., & KOGAN, N. *Modes of thinking in young children: A study of the creativity-intelligence distinction*. New York: Holt, Rhinehart, & Winston, 1965.
- WERTS, C. E. The many faces of intelligence. *Journal of Educational Psychology*, 1967, 58, 198-204.

(Received January 3, 1967)

STUDENT PERCEPTIONS OF COLLEGE PERSISTENCE, AND THEIR INTELLECTIVE, PERSONALITY AND PERFORMANCE CORRELATES

EDMOND MARKS¹

Pennsylvania State University

Student expectancies concerning persistence in college and sources of conflict leading to withdrawal from college were related to selected precollege performance, scholastic ability, and personality variables. Perceived reasons for college withdrawal generated a 3-dimensional space, the defining vectors being academic and work skills and their utilization, motivation, and adjustment. Correlations computed separately for the groups reporting high and low probabilities of college dropout suggested that the former group is more concerned with satisfying the expectancies of their parents and that failure to do so is anxiety and guilt producing. Lack of commitment to educational pursuits coupled with this need to satisfy parental expectancies apparently leads to initially adequate performance but subsequent underachievement. 3 times as many of this group withdraw as do the low probables. Similar differences were observed between remaining and dropout students after 3 terms.

Periodically, the problem of the college dropout emerges in the research and educational literature as one of intense concern, particularly to those who view withdrawal as a measure of institutional efficiency. Despite the effort, both conceptual and research, and the unanimity concerning the wastefulness of attrition (Dressel, 1943; Macintosh, 1948), the college attrition rate remains disconcertingly stable. Summerskill (1962), after reviewing some 35 studies spanning a period of 49 years, concluded that, on the average, 50% of matriculating students withdraw during the normal 4-year period. Data reported by the Dean of Admissions and Registrar at the Pennsylvania State University indicate that 50% of the students enrolling in the fall semester of

1955 withdrew through 5 years. In a more recent study at Penn State, Lindsay, Marks, and Hamel (1966) reported an all-university, that is both main and Commonwealth campus, attrition of 55% within 4 years.

Although estimates of the magnitude of college dropout seem unambiguous, the reasons for and nature of this phenomenon are certainly much less clear. With few exceptions the research on college dropouts is characterized by a lack of an adequate conceptual base and a reliance upon ex post facto methodology. The typical research design employed in this area consists of identifying dropouts and then examining either a broad range of precollege characteristics, for example, age (Gable, 1957), sex (Iffert, 1957), parental occupation (Suddarth, 1957), community size and location (Strang, 1937), scholastic ability (Summerskill, 1962), and high school preparation (Iffert, 1957), or postwithdrawal factors, such as reasons for withdrawal (Iffert, 1957), feelings relating to the educational

¹The author gratefully acknowledges the assistance of Craig Messersmith in the construction of the inventory scales and early development of the study. The author is now at the Georgia Institute of Technology.

experience and withdrawal (Ifert, 1957), motivation (Mercer, 1941; Woods & Chase, 1937), and adjustment (Farnsworth, Funkenstein, & Wedge, 1955). The use of adequate controls necessitated by this type of research has been singularly lacking, which renders the frequently contradictory results only more uninterpretable. The only reliable conclusion emerging from this mass of research activity is that students with poor high school preparation or lower scholastic aptitude (and in many cases, both) have a higher incidence of college withdrawal. Summerskill (1962) provides an excellent review and criticism of the college dropout literature.

An immediate and appealing conjecture is that the majority of withdrawals from college are directly traceable to academic difficulty. What evidence there is, and admittedly this is a difficult thing to measure, suggests that this is not the case. Summerskill (1962) concludes that only about one-third of college dropouts are due to academic difficulty. Of the total number of students matriculating at Penn State in the fall semester of 1955, 15% (or roughly 30% of the dropouts) were dismissed by the University for academic reasons. Data on file with Student Affairs Research at Penn State are also relevant. These data indicate that the correlation between academic achievement, as measured by grade-point average (GPA), and persistence, as measured by the number of terms a student remains in school, is at best moderate, with many withdrawing students (particularly female) having cumulative averages well above the minimum required.

In some cases the reasons for college withdrawal are clearcut, perhaps academic failure or simply running out of funds. If our estimates are any-

where near correct there still remains a sizeable proportion of college dropouts, around 60%, for whom the dropout process is considerably more complex and probably multiply caused. It is this group to which the present studies are addressed.

In the present study college withdrawal is viewed as the behavioral outcome of a decisional process the basic data of which are those perceptions, cognitions, and stimuli which the student judges to be relevant to his college behavior. The typical student entering a college setting has developed through experience with a similar environment, for example, high school, and relationships with significant personages and institutions, such as parents and church, an extensive and intricate set of expectancies relating to his own behavior and the behavior of others, and the consequences of his behavior. These expectancies and their environmental confirmation or disconfirmation serve as antecedents to observed behavior. The student in the two-choice withdrawal situation can be treated as a rational organism who evaluates and relates certain cognitions concerning himself, for example, ability, performance, persistence, educational and career goals, and perceptions and cognitions concerning environmental inputs, for example, parental and peer pressures, educational "payoff," experience of success or failure, and then "decides" on one course or the other. Superimposed on this matrix of perceptions and cognitions and affecting both their formal properties and possibly their relationship to the behavioral outcome, that is, the decision and act of withdrawal, is the class of judgmental skills and preferred modes of dealing with informational input which are broadly termed personality variables.

Within this framework, research efforts would tend to focus on certain perceptual and cognitive structures such as expectancies or anticipations regarding educational behaviors and outcomes, the identification and assessment of a class of admissible end states (goals), the congruence of expectancies and perceived behavioral outcomes, factors relating to the development and expression of expectancies, and finally, changes in cognitive structures and behavior with continued intercourse with the college environment. The influence of theorists such as Bruner (1957), Kelly (1955), Lewin (1936), and Tolman (1951) upon the present development is apparent.

The data reported here are those from the first in a series of studies examining student expectancies concerning withdrawal from college, their precollege ability, personality, and learning correlates, and changes in these expectancies attributable to selected situational inputs, for example, conditions of success or failure, college environments, and so forth. It is intended that these studies extend over the 4-year college experience and include postcollege behaviors, for example, occupational stability and job satisfaction, as well. Examination of expectancies as expressed immediately prior to college, including their precollege ability and personality correlates, and their change or resistance to change as a function of exposure to college may provide a fruitful approach to the study of college dropout.

METHOD

Construction of the Attrition Expectancy Questionnaire

To determine those reasons and areas of conflict which students judge as relevant to

college dropout, a sample of 1000 freshmen entering Penn State in the summer term of 1965 were administered a two-item, free response questionnaire. For both items the subjects (Ss) were told that 50% of the students who matriculate at Penn State fail to graduate. Item 1 simply asked S to list all "your reasons," if he were among the 50% who withdrew. Item 2 asked S to think of a person whom he knew well who would probably be in the dropout group and list his reasons for withdrawing. The items were counterbalanced in their presentation.

A content analysis of Ss' responses yielded some 33 relatively frequent and distinct perceived reasons for college withdrawal, many of which a sophisticated investigator would consider rather naive, for example, "can't cope with pressures," or "dislike the college." These were precisely the "naive" verbalizations in which the authors were interested, however.

These 33 statements of reasons for withdrawing from college, together with 10 statements relating to actual withdrawal or transfer from college formed the Attrition Expectancy Questionnaire (AEQ). Each item was accompanied by a five-point subjective probability scale ranging from highly probable to highly improbable. The questionnaire was introduced by telling Ss that roughly one-half of Penn State students fail to complete their program of study. The S was to respond to each item in terms of how probable that statement was with respect to his college behavior.

Two sample items are:

1. How likely is it that you might interrupt your college program?

1	2	3	4	5
highly				highly
improbable	improbable	neutral	probable	probable

2. If I withdraw it will be due to my poor high school preparation.

1	2	3	4	5
highly				highly
improbable	improbable	neutral	probable	probable

Ability, Personality, and Learning Correlates

Having focused attention upon student expectancies, a legitimate question to ask is how these expectancies are shaped. Most investigators of college dropouts agree that one important factor involves the broad area of family dynamics, particularly parental attitudes and behaviors (Brown,

1963). Some data are available which suggest that certain parental attitudes and behaviors as perceived by the student are related to the student's college performance (Malloy, 1954; Sexton, 1965; Trent, Athey, & Craise, 1965). For the present study, two scales reflecting two important aspects of the parent-student relationship were constructed. The first, parental attitude, assessed the degree of parental acceptance or rejection the student experienced, while the second, parental press, evaluated the perceived influence the parents exerted upon the student's behavior.

The effects of family dynamics as well as other socializing forces, for example, cultural values, peer behaviors, etc., can be extended to more specific cognitive structures and motivational states. Three such areas felt to be important were fear of failure, level of aspiration, and educational values. Again, there is some, though slight, evidence that these variables are related to academic performance (Hancock & Teevan, 1964; Moulton, 1965; Munger, 1956). Three true-false scales were constructed for each of these variables. Together, the five scales were intended to cover a significant portion of the range of parental and social influences upon the students' educational expectancies.

Two other classes of variables important to the formation of these expectancies were also included in the analysis. Intellectual functioning would seem to be relevant in two ways. First, the brighter student simply has a better chance of meeting the academic requirements of his college program. Second, and more subtle, the brighter student would presumably have the skills needed for a more effective and elegant management of his interpersonal and other environmental relationships. As such, one might expect a positive correlation between a measure of scholastic ability and expectancies of college achievement and success.

The set of environmental outcomes, particularly those relating to the student's academic performance, is the second set of interest. During high school, the student has experienced certain academic and career-relevant feedback which falls broadly under the rubric of success-failure. This environmental confirmation or disconfirmation would serve either—in the former case—to strengthen expectancies, or to induce cognitive realignment or change in the latter case (Bruner, 1957). A measure of high school academic achievement would provide some indication of this feedback.

Procedure

The 43-item AEQ and the five scales reflecting level of aspiration, fear of failure, educational values, and parental attitudes and press were administered to a random sample of Penn State freshmen immediately prior to their registration for the first term of work. In addition to these measures, the SAT verbal, quantitative, and total scores, high school average, and a measure of high school quality—the percentage of students in the high school graduating class going to college—were obtained for each student. This latter variable was included to evaluate and control for possible differences in quality among the high schools from which the *Ss* came (Marks & Murray, 1965).

The first term and third term cumulative GPAs as well as the dichotomized score—either 0 or 1—indicating withdrawal or nonwithdrawal from the university after 1 year of study were included.

Several related analyses were undertaken on these data. First, the 43 AEQ items were intercorrelated and the pattern of correlations examined for a smaller number of explanatory dimensions of college dropout. It was originally intended to factor this correlation matrix, but upon examination the joint probability density function of these variables was so obviously nonnormal that the results of a factor analysis would be highly questionable. As a result, "factoring" was done by inspection.

Second, and again using the entire sample, the correlations between the 43 AEQ items and the other 13 variables as well as the intercorrelations among these latter variables were computed and examined.

To determine differences in the frames of reference between those *Ss* reporting a high probability of withdrawing from college and those reporting a low probability, means, standard deviations, and covariances of the 43 AEQ items and 13 variables were computed separately for the two groups. Special interest was directed toward possible differences between corresponding correlation values.

Finally, the sample was again broken down on the criterion—withdrew from or remained in college—and the means, standard deviations, and covariances of all variables computed for the two groups.

The reason for examining covariance matrices separately for both the attrition and nonattrition groups and the probable and nonprobable attrition groups was to uncover any possible nonadditive effects re-

lated to the two variables, expected dropout and actual dropout.

Subjects

The Ss were 300 students selected randomly from the entire freshman class matriculating at Penn State in the fall term of 1965. There were 229 males and 71 females. All were tested prior to registration for their first term, so that none had experience with the college academic environment and little, if any, contact with its social environment.

RESULTS

An incidental but interesting finding concerned the attribution of dropout as elicited by the two initial free response items. When referring to themselves Ss spoke mostly of external or personally acceptable causes, e.g., poor preparation, lack of funds, present school not first choice. Reasons given for the dropout of another person were imputed almost entirely to personal weaknesses, e.g., immaturity, maladjustment, lack of self-discipline. Apparently, certain defensive strategies are operative in the responsivity to such items, a finding which has implications for the post-withdrawal type of attrition questionnaire.

The results of the crude clustering of the correlations among the 43 AEQ items—crude in the sense that a precise analytic procedure was lacking—suggested three major dimensions accounting for the correlations. For this clustering only correlations greater than $|.30|$ were used. One dimension of perceived college withdrawal involved academic skills and utilization of such skills. Items "loading" high on this dimension were lack of ability, poor study habits, inability to concentrate, poor high school preparation, and laziness. A second rather clear-cut dimension related to motivation. Describing this dimension were such items as want to do other things, prefer to "bum around," no

reason for attending college, lack of occupational or career goals, and again laziness. The third and most extensive dimension involved what was simply called adjustment. Items defining this dimension were inability to handle independence (the main campus of Penn State is attended almost entirely by resident students), immaturity, homesickness, inability to cope with academic and social pressures, can't get along with other students, academic frustration, disillusionment, inability to conform, and wanting to go to a smaller school.

These dimensions did not appear to be completely distinct, there being some slight overlap of items. In general, however, the three areas seemed clearly defined. Two minor causes of dropout which also emerged from this analysis were social or nonacademic overinvolvement, and leaving for a job or marriage.

The correlations between the 43 AEQ items and the 13 ability, personality, and performance variables were generally rather low (although many reached the r value of $.14$ needed for significance at the $.01$ level), except for the level of aspiration, parental attitude, and fear of failure scales.

Students who reported a high level of aspiration felt it was highly *unlikely* that they would encounter difficulty or leave school because of lack of occupational or career goals, poor motivation, lack of self-discipline, inability to concentrate, poor study habits, lack of ability, laziness, or a general unreadiness for college. In addition, such students reported that it was highly *unlikely* that they would interrupt their college studies or transfer to another school.

On the other hand, students who reported a high fear of failure felt that if they did drop out of college it would

likely be due to an inability to concentrate, inadequate high school preparation, inability to cope with academic pressures or teaching, academic frustration, and lack of occupational and career goals.

Students who reported their parents as accepting and supporting felt it was *unlikely* that they would interrupt their college career because of inadequate ability, poor study habits, lack of self-discipline, insufficient motivation or career goals, or laziness.

Two other correlations are also worth noting. Neither the brighter students (SAT) nor the higher-achieving students (GPA) felt that inadequate high school preparation was a probable source of college withdrawal.

For all comparisons a correlation coefficient of $|.20|$ or higher was required. Correlations of the AEQ items with withdrawal-nonwithdrawal are reported separately in the section relating to actual dropout.

Differences between Probable and Nonprobable Dropouts

Of the 300 students examined, 35 reported that it was highly probable or probable that they would discontinue their studies at Penn State. Ninety-seven students reported that it was highly improbable that they would discontinue.

As might be expected, there was a significant difference between the mean vectors of the remaining 42 AEQ items for these two groups, the T^2 value being significant at the .05 level. For every item, the probable group reported a higher probability of withdrawing for the stated reason. Those items most clearly differentiating the two groups covered only two of the three dimensions previously mentioned, including specifically poor motivation, inability to work to potential, laziness, frustra-

tion, lack of occupational or career goals, wanting to "bum around," external attractors such as employment, marriage, or the Army, etc., immaturity, not knowing "why I'm here," and a general inability to adjust to or cope with academic pressures. The area where differences were much smaller, if not negligible, was that of academic and work skills and their utilization.

The means, standard deviations, and intercorrelations among the 11 (third term cumulative GPA was deleted because of numerous zero values) ability, personality, and performance variables are reported separately for the two groups in Table 1.

Only three of the mean differences were significant at the .05 level. Students assigning a low probability to their possible withdrawal from college obtained higher level of aspiration and educational values mean scores than those students assigning a high probability to this behavior. Most interesting, the nonprobable dropout group also had a significantly higher mean withdrawal score. This value reflects the number of dropouts—the higher the value the fewer dropouts. Transforming these means to percentages indicated that 30% of those students reporting probable dropout withdrew within three terms whereas only 10% of the nonprobable group withdrew.

Examination of the correlations indicated that the brighter students, as usual, performed better in college, performed better in high school, and attended the better high schools. In addition, they tended to report low fear of failure. The higher-achieving students, aside from being brighter and having better high school records, also reported their parents as being more accepting and supporting than

TABLE 1
MEANS, STANDARD DEVIATIONS AND INTERCORRELATIONS* AMONG THE ELEVEN ABILITY, PERSONALITY, AND PERFORMANCE VARIABLES FOR THE PROBABLE AND NONPROBABLE DROPOUT GROUPS

	SAT V	SAT Q	SAT T	1st term GPA	Fear of failure	Educational values	Level of Aspiration	Parental attitude	Parental press	HSA	% College
SAT Q	23										
SAT T	(30)	(79)									
1st Term GPA	(75)	(15)	(32)								
Fear of failure	(34)	(38)	(-27)	(-19)							
Educational values	(-04)	(-13)	(-01)	(-02)	(-17)						
Level of aspiration	(-15)	(14)	(-27)	(-02)	(-16)	(41)					
Parental attitudes	(-27)	(14)	(-27)	(-02)	(-16)	(07)	(04)				
Parental press	(27)	(-11)	(08)	(-36)	(-02)	(15)	(18)	(-20)			
HSA	(-40)	(19)	(-13)	(-05)	(-07)	(15)	(00)	(-08)	(04)		
% College	(14)	(05)	(09)	(32)	(-20)	(02)	(14)	(-05)	(12)	(-20)	(38)
$\bar{X}$	(-02)	(20)	(15)	(19)	(-07)	(29)	(-03)	(7.9)	(4.9)	(2.91)	(30)
SD	(562)	(578)	(1143)	(2.52)	(3.2)	(7.5)	(12.0)	(1.9)	(1.5)	(.54)	(19)
	(72)	(82)	(124)	(.90)	(1.7)	(2.4)	(3.3)		(2.1)		(17)

Note.—The values for the probable dropout group are given in parentheses. For the probable group, $N = 35$; for the nonprobable dropout group, $N = 97$.
* All decimal points for intercorrelations have been omitted.

did the lower-achieving students. High school average and parental attitudes were also positively correlated. Students who attended the better high schools tended to report higher educational values and lower fear of failure, while parental attitude, already shown to be related to high school and college achievement, was negatively correlated with perceived parental press.

Although a thorough examination of the nonadditive effects of probable or nonprobable withdrawal would require testing the equality of the complete variance-covariance matrices of the two groups (Anderson, 1958), only a few of the pairs of correlations were of interest, in particular those involving GPA and the personality variables. Tests of significance were made on selected pairs of correlation coefficients by use of the test criterion $W = \sum (n_i - 3) (Z_i - \bar{Z})^2$ (Graybill, 1961).² Those correlations whose differences were significant at the .05 level are underlined in Table 1.

The negative correlation between fear of failure and GPA, and the positive correlation between educational values and level of aspiration for the probable dropout group are to be compared with the essentially zero correlations between these variables for the nonprobable group. Most striking is the positive correlation between fear of failure and parental attitudes for the probable group, while this correlation is negative for the nonprobable dropouts.

Differences between Dropouts and Nondropouts

Of the 300 students studied, 39, or 13%, withdrew during the first three terms. The remaining 261 students completed at least three terms of study.

² $Z = \text{arctanh } r$.

TABLE 2
MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS^a AMONG THE TWELVE ABILITY, PERSONALITY, AND PERFORMANCE VARIABLES FOR THE DROPOUT AND REMAINING GROUPS

	SAT V	SAT Q	SAT T	GPA (1)	Fear of failure	Educational values	Level of aspiration	Parental attitude	Parental press	HSA	% College	GPA (3)
SAT Q	(02) 24											
SAT T	(76) 76	(66) 80										
GPA (1)	(19) 20	(06) 15	(19) 24									
Fear of failure	(24) -14	(-05) -32	(15) -29	(16) 09								
Educational values	(-25) -05	(-01) 16	(-20) 07	(-05) 09	(21) -17							
Level of aspiration	(-27) -19	(-04) 11	(-23) -06	(-06) -09	(-36) -14	(35) 22						
Parental attitude	(-01) 04	(04) -10	(02) -04	(27) 20	(-02) -02	(25) 09	(-32) 18					
Parental press	(-24) -17	(18) 05	(-06) -07	(-03) -07	(21) 20	(28) -08	(23) 03	(01) -44				
HSA	(21) 20	(27) 19	(33) 25	(32) 35	(08) 01	(-25) 05	(00) 10	(08) 26	(35) -06			
% Coll.	(09) 09	(14) 18	(16) 17	(-04) 03	(10) -26	(-14) 26	(-34) 03	(42) 05	(17) -12	(01) -07		
GPA (3)	22	16	25	78	06	08	00	24	24	47	-01	
SD	(543) 557	(587) 599	(1110) 1157	(1.86) 2.83	(3.6) 2.8	(7.6) 8.2	(12.7) 14.6	(6.90) 8.0	(4.8) 5.1	(2.73) 3.03	(34) 39	(0.0) 2.68
X	(87) 71	(75) 80	(116) 119	(1.07) .71	(1.6) 1.6	(2.2) 1.8	(3.6) 3.4	(1.50) 1.9	(1.8) 2.0	(.60) .45	(13) 18	(.00) .56

Note.—The values for the dropout group are given in parentheses. For the dropout group, $N = 39$; for the remaining group, $N = 201$.

^a All decimal points for intercorrelations have been omitted.

The most striking differences between the two groups on the 43 AEQ items involved the perceived chances of discontinuing studies and, second, transferring to another school. In both cases, students who withdrew assigned a higher probability to those two outcomes than did the students who were still enrolled. Again, the mean vectors were significantly different. Those items most clearly differentiating the two groups were poor high school preparation, lack of occupational or career goals, and six items relating to adjustment. As with the probable-nonprobable dropout dimension, withdrawal due to academic ability and skills and their utilization was not considered relevant. It should be recalled, however, that the probable and actual dropout groups were not independent, roughly 30% of the former actually withdrawing from college.

The means, standard deviations, and intercorrelations among the 12 ability, personality, and performance variables are reported separately for the two attrition groups in Table 2.

There were several significant mean differences. Students who remained in college had higher SAT scores, higher high school averages, higher first term GPA's, and also had higher level of aspiration and parental attitude mean scores. On the other hand, dropouts reported higher feelings of fear of failure.

Nonadditive effects of college dropout were again tested by examining selected pairs of correlation values. Significant differences are underlined in Table 2. Students withdrawing from college and reporting a high fear of failure reported higher educational values and tended to come from the better high schools. For those students remaining in college these correlations were both negative. Again, for the withdrawing students the correlations

between level of aspiration and parental attitudes, and level of aspiration and percent going to college were negative, while for the remaining students these values were low and positive. Although high parental press for the dropout group was accompanied by high educational values and high high school average, these correlations were low and negative for the remaining students. There was a high negative correlation between parental attitude and press for the remaining students, while for the dropouts this value was zero.

DISCUSSION

It is apparent from the present data that students arrive at college with certain rather well-formed perceptions and cognitions relating to their college attrition behavior, and that these expectancies, as hypothesized, have selected precollege and postcollege admission correlates. In addition, these expectancies show considerable individual differences.

At a general level, the student in the post-high school—precollege interval rather simply perceives the sources of college dropout as falling in a three-dimensional space whose defining vectors were labeled, (a) academic and work skills and their utilization, (b) motivation, and (c) adjustment or coping behaviors. In describing their own behavior, however, few students were willing to identify themselves with the first vector, and probably with good cause since most colleges and universities are unwilling to accept students of low scholastic ability.

At the same level it was indicated that the strength and direction of expectancies involving dropout and its antecedents were related to level of aspiration, fear of failure, and parental attitude. Without getting involved in questions of cause and ef-

fect, these correlations suggest that the student maintains a balance (Heider, 1958) or congruence (Festinger, 1957) among the elements of his cognitive and motivational domains. As such, attrition expectancies can, in part, be viewed as a cognitive matching between salient motivational states, for example, fear of failure, etc., and relevant situational outcomes, such as academic performance.

The role of parental figures both in the development of educational expectancies and in student academic achievement (behavior) cannot be underestimated. Clearly, the environment established by the parents and the student's involvement with this environment affects the student's perceptions of and characteristic modes of dealing with extra-family environments. Higher high school and college achievement, despite an absence of correlation with ability, attest to the facilitating properties of parental attitudes.

Quite unexpected was the lack of relationship between attrition expectancies and scholastic ability and previous educational experiences, for example, high school average and high school quality. What did emerge, however, was a pattern of correlations among the ability, personality, and performance variables which was suggestive of a reciprocal reinforcement by the parental and school environments. Students' educational values are apparently mutually reinforced and shaped by the joint efforts of parents and school; a not so surprising notion which offers the interesting research possibility of studying discrepancies between these two behavioral domains.

Turning to differences between students reporting high and low probabilities of college dropout, it was indicated that the probable group obtained lower level of aspiration

and educational values mean scores. These students are apparently less committed to education and obtaining a degree than are the improbable dropout group. More interesting, however, are those correlations described previously which are significantly different for the two groups. This pattern of correlations suggests that the probable dropout group is much more concerned about satisfying the expectancies of their parents and that failure to do so might be anxiety and guilt producing. For those members of this group who view their parents favorably, that is, high parental attitudes and low press, the conflict arises in relation to their lack of commitment to educational pursuits. The motivational properties associated with this conflict state apparently lead to an initially adequate academic performance and perhaps even overachievement. It is questionable whether resolution of this conflict would continue to have this effect upon performance, in fact, the mean third term cumulative GPA of the remaining students in this group was considerably lower than that of the corresponding nonprobable dropout group mean. Those students in the probable dropout group who report low parental attitudes and high parental press are among the earliest poor achievers and dropouts.

In terms of the prediction of college dropout, there is a significant relationship between a student's expectancy of, and his actual attrition behavior. As to magnitude, fully three times as many students in the probable dropout group withdrew through three terms as did students in the nonprobable group. If you want to know whether a student is a potential college dropout, a good starting place is simply to ask him.

The reasons why he says "yes" are indeed more complex.

The final aspect of the present study was a comparison of the dropout and remaining students after three terms of work. As frequently recorded, the dropout demonstrates both lower ability and poorer high school performance. In addition, the dropout group had a lower mean first term GPA, this value being below the minimum required for graduation from the university. As with the probable-nonprobable dropout dimension, level of aspiration was significantly related to withdrawal. Parental attitudes continued to be related to performance criteria, being positively correlated with GPA and persistence. A higher fear of failure mean score was also observed for the dropout group. These data complement those of Pervin and Rubin (1965) relating to discrepancies between self and environment and subsequent dropout behavior.

Together with the pattern of significantly different correlations obtained here, the data from these studies suggest that students dropping out of college have difficulty in resolving conflicts involving commitment to educational behaviors, drives toward parental and career satisfaction, and discrepancies between cognitions concerning self and environmental feedback. That these students are aware of the existence of these conflicts and their possible outcomes—although they may be unaware of the dynamics involved—is indicated by the dropout students' precollege expectancies.

Further studies in this series will examine changes in precollege expectancies as a result of environmental interaction, and their relation to long range criteria related to educational and career success.

REFERENCES

- ANDERSON, T. W. *An introduction to multivariate statistical analysis*. New York: Wiley, 1958.
- BROWN, D. R. Personality, college environment and academic productivity. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1963. Pp. 536-562.
- BRUNER, J. S. On perceptual readiness. *Psychological Review*, 1957, **64**, 340-358.
- DRESSEL, P. C. Liberal arts students advised to withdraw. *Journal of Higher Education*, 1943, **14**, 43-45.
- FARNSWORTH, D. S., FUNKENSTEIN, D., & WEDGE, B. A study of the social and emotional adjustment of the "early admission" college student. Report for the Fund for Advancement of Education, 1955. (Mimeo)
- FESTINGER, L. *A theory of cognitive dissonance*. Stanford: Stanford University Press, 1957.
- GABLE, R. I. A study of the student dropout problem at Miami University. *Dissertation Abstracts*, 1957, **17**, 61.
- GRAYBILL, F. A. *An introduction to linear statistical models*. Vol. 1. New York: McGraw-Hill, 1961.
- HANCOCK, J. G., & TEEVAN, R. C. Fear of failure and risk-taking behavior. *Journal of Personality*, 1964, **32**, 200-209.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- IFFERT, R. E. Retention and withdrawal of college students. United States Department of Health, Education and Welfare, Bulletin 1958, No. 1. Washington: United States Government Printing Office, 1957.
- KELLY, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
- LEWIN, K. *Principles of topological psychology*. New York: McGraw, 1936.
- LINDSAY, C. A., MARKS, E., & HAMEL, L. S. Native and transfer baccalaureate students. *Journal of College Student Personnel*, 1966, **7**, 5-13.
- MACINTOSH, A. *Behind the academic curtain*. New York: Harper, 1948.
- MALLOY, J. An investigation of scholastic over- and under-achievement among female college freshmen. *Journal of Counseling Psychology*, 1954, **1**, 260-263.
- MARKS, E., & MURRAY, J. E. Nonadditive effects in the prediction of academic achievement. *Educational and Psychological Measurement*, 1965, **25**, 1097-1104.
- MERCER, M. A study of student mortality in a home economics college. *Journal of Educational Research*, 1941, **34**, 531-537.

- MOULTON, R. W. Effects of success and failure on level of aspiration as related to achievement motives. *Journal of Personality and Social Psychology*, 1965, **1**, 402-406.
- MUNGER, P. Student persistence in college. *Personnel and Guidance Journal*, 1956, **35**, 241-243.
- PERVIN, L. A., & RUBIN, D. B. Student dissatisfaction with college and the college dropout: A transactional approach. Princeton University, 1965. (Mimeo)
- SEXTON, V. S. Factors contributing to attrition in college populations: Twenty-five years of research. *Journal of General Psychology*, 1965, **72**, 301-326.
- STRANG, R. *Behavior and background of students in college and secondary school*. New York: Harper, 1937.
- SUDDARTH, B. M. Factors influencing the successful graduation of freshmen who enroll at Purdue University. Progress Report No. 21, 1957, Purdue University. (Mimeo)
- SUMMERSKILL, J. Dropouts from college. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1962. Pp. 627-657.
- TOLMAN, E. C. A psychological model. In T. Parsons & E. A. Shils (Eds.), *Toward a general theory of action*. Cambridge, Mass.: Harvard University Press, 1951. Pp. 279-361.
- TRENT, J. W., ATHEY, I. J., & CRAISE, J. L. Technology, education, and human development. *Educational Record*, 1965, **46**, 93-103.
- WOODS, A. H., & CHASE, G. Forms of personality obstructive to progress in college. *Journal of Social Psychology*, 1937, **8**, 411-431.

(Received July 28, 1966)

LEARNING OF LOGICAL CONNECTIVES BY ADOLESCENTS WITH SINGLE AND MULTIPLE INSTANCES¹

JAMES YOUNISS AND HANS G. FURTH

*Center for Research in Thinking and Language,
Catholic University of America*

Adolescents classified with correction a set either of single or of multiple instances according to 3 logical concepts: conjunction, exclusive disjunction, and conjunction with negated elements. Responses were symbolic expressions composed of alphabet symbols standing for elements and other symbols standing for logical connectives. On a subsequent noncorrected criterial task new instances of these concepts were presented as well as instances of novel problems of logical addition and simplification. No quantitative difference was observed in the attainment series between Ss classifying the single or the multiple sets. Ss who classified multiple instances initially, however, showed greater success in the criterion task including performance on the novel problem of addition. Also, Ss in the multiple instance condition showed patterns of success distinct from those evidenced by Ss dealing with a single instance set. It was concluded that training with multiple instances enhanced rule learning of the logical connectives and reduced point-to-point matching of instances with symbols.

The present study was designed to assess adolescents' attainment of combinatorial rules defined by the logical connective—"and," "not," and "or"—and to investigate the degree of mastery of these logical expressions. Following the general procedure of Youniss and Furth (1964), a concept acquisition and subsequent criterial series were constructed in which adolescents had to match instances to symbolic expressions. Since elements of the instances were defined beforehand, subjects (Ss) had only to attain the meaning of the logical connectives within each symbolic expression. This procedure follows a rule-learning paradigm (cf. Haygood & Bourne, 1965) rather

than an "attribute discovery" procedure in which Ss must discover relevant elements as well as the combinatorial rules. Two major modifications of the previous study were: (a) a comparison of attainment with single and with multiple instances of correctness and (b) an extension of criterion trials to include novel types of problems not seen during acquisition but which could have been learned implicitly during the acquisition series.

The first major analysis in the present study compared performance between Ss classifying each concept according to one particular instance with performance in which each concept was classified according to at least two distinct instances. In the condition with the single concept instance for example, only one instance showing two elements was ever correct for conjunction. With multiple correct instances, more than one instance, each showing a different pair of elements, was correct for conjunc-

¹ This research was supported in part by Vocational Rehabilitation Administration Grant RD-1484 and by National Institute of Child Health and Human Development Grant HD-02026-01. Thanks are given to Margaret Meade who collected the data and to school administrators and teachers who provided their cooperation.

tion. To illustrate, in the single-instance condition, the one correct instance for conjunction showed the colors red and blue together, while in the multiple-instance condition the instance showing red and blue and the instance showing a man and a house together were both correct for conjunction.

This comparison of type of instance presentation was suggested by the observation that college students demonstrated greater concept-generalization after training with varied instances than after training with uniform instances (Turnure & Wallach, 1962). Since in the present study the number of exposures for matching instances with concepts was held constant, presentation of varied instances necessarily entailed a reduction in the number of trials in which any particular instance was matched with a concept. On the other hand, showing one particular correct instance repeatedly might result in more stable acquisition. Thus, it was not obvious that results similar to Turnure and Wallach's would be obtained with adolescents.

The second modification of our previous study was designed to determine whether Ss implicitly learned the logical rules for addition and simplification during acquisition. This question could be answered by comparing performance with instances essentially similar to those presented in acquisition against performance with novel instances showing fewer elements (addition) or more elements (simplification) than were necessary to exemplify correctly a concept.

This comparison was suggested by results with adolescents (Furth & Youniss, 1965) showing that ease of classification of some logical rules is contingent upon the physical compo-

sition of instances. For example, it was found that for the exclusive disjunctive rule, "either X or Y," instances showing a single element were easiest to classify, instances showing one relevant and one irrelevant element were next easiest, and instances comprised of two irrelevant elements (denial) were hardest to classify. This result was consistent among five independent groups of Ss and suggests that adolescents do not necessarily learn logical rules apart from their concrete application. In the present design, evidence for rule learning would be obtained if Ss showed success with the problems of addition and simplification that presented novel instances but followed the same rules for connectives as were learned during acquisition.

METHOD

General Design

An acquisition series of 48 corrected trials presented 16 instances each of conjunctive (C), exclusive disjunctive (D), and conjunction with negation (N) concepts. The Ss were exposed either to unitary instances, called *single-instance* condition or to two kinds of instances for each concept, called *multiple-instance* condition. The single-instance condition presented to approximately one-half the Ss instances comprised of two colors and to the other one-half instances consisting of two familiar figures. The condition of multiple instances presented alternately color and figure instances. Each instance was to be matched to one of three symbolic expressions— $X \cdot Y$, X/Y , and $\bar{X} \cdot \bar{Y}$ —which stood for C, D, and N concepts, respectively. The Ss were told which alphabet-symbols stood for which elements on the instance cards and that they were to learn the meaning of the symbolic expressions which stood for the connectives.

Following the acquisition series 20 non-corrected criterial trials were given; four trials each which instanced C, D, and N concepts; and eight more trials which instanced the disjunctive class in new ways. All Ss performed the criterial series with instances of geometric forms which were not shown in the acquisition series. Again Ss

were told the meaning of the alphabet-symbols, but since correction was not given a situation was provided to test whether Ss had learned and could apply the logical connectives to which they had previously been exposed.

Procedure

The Ss were tested in their classrooms in groups of approximately 40 with the classroom teachers present serving as monitors. Each S was given a two-page handout. Sheet 1, for the *attainment series*, contained 48 trial markers; after each trial the three symbolic expressions were given. As these were being passed out, the experimenter (E) put on the blackboard the same symbolic expressions which appeared on the sheet as well as the elements of the instance cards which the alphabet-symbols represented.

Single-instance condition with colors. Instances used in this task were two color patches separated left-from-right on 5×8 inch cards. Sixteen of these cards contained the colors red and blue; these cards instantiated the C concept: "red and blue colors." Sixteen other cards contained red and yellow, red and green, blue and yellow, or blue and green colors. These cards were correct instances of the D concept: "either red or blue, but not both of these colors." The remaining 16 cards were colored yellow and green, yellow and orange, green and yellow, or green and orange; these were instances of the N concept: "neither red nor blue."

On the answer sheet next to each trial number appeared the response alternatives $R \cdot B$, R / B , and $\bar{R} \cdot \bar{B}$, the symbolic expressions for C, D, and N, respectively. Left, middle, and right position of each alternative was counterbalanced over the attainment series in this and in each of the other conditions described below.

The Ss were told beforehand that "R" stood for red and that "B" stood for blue. Further, they were told to learn which "answer" (i.e., symbolic expression) was the correct one for each card they would be shown. They were informed also that E would provide the correct answer for each card after everyone had circled his choice and raised his pencil. Correction was carried out by E's pointing to the correct expression of the three placed on the blackboard.

In this and all other conditions instances were presented singly in a restricted random order so that within six trial blocks two instances of each concept were presented. A

random order following this restriction was designed for 48 trials.

Single instance condition with figures. This condition was similar to that described above except that figures replaced colors with an appropriate change in alphabet symbols. The C concept in this condition was "man and house," which was expressed by $M \cdot H$. "Either man or house, but not both" was the D concept which was expressed M / H . Finally the N concept of "Neither man nor house," expressed $\bar{M} \cdot \bar{H}$, completed this condition of 48 trials with 16 trials per concept.

Multiple instance condition. This condition, also consisting of 48 trials, presented both color and figure instances and required both types of symbolic responses. The 16 trials illustrating each concept were divided equally between color and figure instances. Odd-numbered trials presented color instances and even-numbered trials presented figure instances. The same random order prevailed as was used above. In all other respects this condition resembled the single-instance conditions.

Immediately following Trial 48 of the attainment series Ss were told to turn to Sheet 2 of their handout, beginning the *critical series*. All Ss were then exposed to 20 noncorrected trials. In this series geometric forms were used as instances and appropriate alphabet symbols along with the connectives used in attainment were responses.

Trials 1 to 12 of this series showed in random order four instances of C, D, and N concepts. For the C concept, the instance card showed a circle and a square; the appropriate expression was $C \cdot S$. The correct instances for the D concept presented a circle and triangle or a square and triangle which was expressed as C / S . The four N instances showed a triangle and ellipse, expressed as $\bar{C} \cdot \bar{S}$. The Ss were told beforehand that C stood for circle and S stood for square.

The next four trials, called affirmation with simplification (AS), presented also dual-element instances comprised of geometric forms. They differed from all other trials in this experiment in that they instantiated affirmation of a unitary class expressed by a single alphabet symbol. Of the two elements comprising each instance, one instantiated the class while the other was irrelevant. To illustrate, Trial 16 showed the instance *square and triangle* with the response alternatives C, S, and $\bar{S} \cdot \bar{T}$. The

correct response here was S (square); a unitary class instanced by the presence of a square and, incidentally, the presence of the irrelevant triangle.

The last four trials of the noncorrected criterial series presented instances of exclusive disjunctive concepts which differed from D trials described above in that only single element instances were shown. We call these trials cases of disjunction with addition (DA). On each of these trials two D-type expressions were alternatives. To illustrate, the last trials of the series showed as the instance a *circle* with the response choices C, S / T, and C / S—the last being the correct response. Thus, in distinction from D and from AS trials only one element comprised an instance but like D and AS problems only one element, in fact, was necessary to satisfy the symbolic expression.

Sampling and Subjects

The adolescents selected for this experiment were taken from an initial sample of 454 children enrolled in Grades 6 and 7 of a parochial school in suburban Philadelphia, Pennsylvania. The total sample was approximately halved when Ss were divided into those failing and those succeeding to an arbitrary criterion in the attainment series. This criterion required three successive correct responses on all three concepts consecutively; it could be achieved with from 9 to 12 consecutive correct responses depending on the point at which it was attained in the series.

Approximately equal percentages of Grade 6 and Grade 7 Ss succeeded, 46% and 50%, respectively; hence it was considered appropriate to combine grade levels into a single adolescent group.

Next an attempt was made to obtain equal size samples from conditions of single

and multiple instances. Since 46% of the single-condition Ss and 50% of the multiple-condition Ss succeeded to criterion, it was concluded that neither of these conditions contributed disproportionately to failure rate; hence scores from Ss in these respective conditions were considered comparable.

An additional sampling restriction was designed to control for speed of initial attainment. Thus, of all Ss achieving criterion, equal numbers were selected from those succeeding in approximately the first, second, and third quarters of the attainment series.

The resulting present sample had a total of 96 Ss; 48 of those who had succeeded in single and 48 who had succeeded in multiple conditions. Further the 48 Ss in each condition comprised 16 Ss who had reached criterion between Trials 1 and 12, 16 Ss who attained criterion between Trials 13 and 22, and 16 Ss who succeeded between Trials 23 and 38; the last trial in which an errorless criterion run could begin. These last three groups of Ss are hereafter referred to as fast, intermediate, and slow learners, respectively. Finally, within single-instance conditions one-half of the fast, slow, and intermediate groups was composed of Ss trained with color instances while the other one-half was exposed to figure instances.

RESULTS

Attainment series. Table 1 reports trials-to-criterion as a function of type of instance and speed of learning and errors-to-criterion as a function of these conditions and of concepts. It can be seen that selection according to speed of attainment resulted in three distinct groupings of

TABLE 1
MEANS OF TRIALS- AND ERRORS-TO-CRITERION IN ATTAINMENT AS A FUNCTION OF
EXPERIMENTAL CONDITIONS AND CONCEPTS

Speed of attainment	Single instance				Multiple instance			
	Trials	Errors			Trials	Errors		
		C	D	N		C	D	N
Fast	4.9	.69	1.12	1.00	6.3	.63	1.31	1.19
Intermediate	18.3	2.63	3.44	2.56	17.6	1.25	2.75	3.31
Slow	29.9	3.94	5.63	5.06	30.6	3.69	5.63	5.94

Ss and that trials-to-criterion scores, in each case, are comparable for single- and multiple-instance conditions.

Error scores were used to evaluate relative difficulty among concepts C, D, and N. It is seen in Table 1 that in five of the six independent groups (3 speeds of learning $\times$ 2 types of instance) fewest errors were made with the C concept, while in four of the six groups D proved to be associated with the most errors. In order to take instance conditions into account speed of learning was collapsed and two Friedman two-way analyses were made, one for single and one for multiple conditions. In both cases reliable differences among concepts were observed; the single condition resulted in $\chi^2 = 5.45$, $df = 2$, $p < .05$; and for the multiple condition yielded $\chi^2 = 24.48$, $df = 2$, $p < .001$. It can be seen in Table 1 that these results reflect differential ordering among concepts due to type of instance condition. For single-instance Ss, C was easiest and D the most difficult while for multiple-instance Ss C was easiest but N was the most difficult concept.

To assess further these dissimilar orders of difficulty, Mann-Whitney U tests were computed between single and multiple conditions for each concept. None of these comparisons indicated a significant difference due to mode of condition. Thus, it was concluded that prior to criterion trials no quantitative differences between

single and multiple conditions were found except that these conditions resulted in differential orders of difficulty among concepts.

Criterion series. Table 2 reports error scores for the five concepts tested in this series as a function of the experimental conditions. An initial analysis was performed to determine overall effects of instance conditions and speed of initial learning on criterial performance. Errors were summed across concepts, yielding a possible range of scores from 0 to 20, and a 3×2 analysis of variance was computed. Single versus multiple conditions proved to have a reliable effect with $F = 12.08$, $df = 1, 90$, $p < .001$; and speed of acquisition also was a reliable source of variance with $F = 10.79$, $df = 2, 90$, $p < .001$; and their interaction resulted in a non-reliable $F < 1$. Thus, facility with which criterion was reached in the attainment series continued to show an effect during criterial trials, even though all Ss were exposed to 48 corrected trials. On the other hand, while instance conditions failed to produce significant effects in attainment trials, a clear difference between them is observed in the criterial series with Ss exposed to multiple instances making consistently more accurate identification of new instances than Ss trained in the single condition.

The second analysis of criterial performance measured differential difficulty among concepts. It is seen

TABLE 2
MEAN ERRORS PER CONCEPT IN THE CRITERION SERIES AS A FUNCTION OF CONDITIONS AND CONCEPTS

Speed of attainment	Single instance						Multiple instance					
	C	D	N	AS	DA	collapsed	C	D	N	AS	DA	collapsed
Fast	.5	.8	1.4	2.5	1.6	6.8	.1	.5	.5	1.8	1.1	4.0
Intermediate	.6	1.6	2.1	2.4	2.2	8.9	.2	.8	.9	2.1	1.0	5.0
Slow	1.1	2.0	2.0	3.0	2.6	10.7	.8	1.6	1.7	2.9	2.0	9.0

TABLE 3
PERCENTAGES OF SUBJECTS SUCCEEDING ON EACH CONCEPT

Speed of attainment	Single instance					Multiple instance				
	C	D	N	AS	DA	C	D	N	AS	DA
Fast	88	75	69	25	38	94	88	81	50	56
Intermediate	81	25	31	25	19	94	63	69	31	69
Slow	69	25	38	13	13	75	44	44	19	25
Collapsed	79	42	46	21	23	88	65	65	33	50

in Table 2 that a clear consistency is obtained across the six independent groups with fewest to most mean errors following the order: C, D, N, DA, AS—with two slight exceptions—these two cases showing equivalent errors in D and N.

Because of the stability of this ordering a further attempt was made to measure interconcept differences. With possible errors per concept ranging only from 0 to 4, an arbitrary score of success was employed. Success with a concept was considered when an S made either one or no errors in the four trials in which that concept was presented; two or more errors then was considered failure. Table 3 reports percentages of success as a function of conditions using this arbitrary criterion of success.

With speed of attainment collapsed comparisons of numbers of Ss succeeding were made between single and multiple conditions for each concept. Three of the five comparisons proved to be reliable although with all five concepts greater numbers of Ss succeeded with multiple-instance training than with single instance. For C a reliable $\chi^2 = 4.66$, $df = 1$, $p < .05$, was obtained; for D a reliable $\chi^2 = 4.18$, $df = 1$, $p < .05$, was observed; and for DA a $\chi^2 = 6.47$, $df = 1$, $p < .02$, was obtained; the result for N approached significance with $\chi^2 = 2.70$, $df = 1$, $p < .10$.

In order to answer the further

question whether adolescents succeed by matching elements of instances with elements of symbolic expressions or whether they succeed by applying rules, performance was measured according to success on all three of the original concepts. While C and N allowed for point-to-point matching, D did not, since it was always instanced by the presence of one relevant and one irrelevant element. That is to say, one could succeed on C or N by merely matching each symbol in an expression with the two elements of an instance without taking into account the conjunctive connection. For D, on the other hand, the connective had to be applied since one of the two symbols had no counterpart in either of the elements in an instance. Thus, if an S succeeded with all three concepts then it would follow that matching alone would not account for his performance.

Among Ss trained in the single condition 35% succeeded on all three concepts while among those in the multiple condition 56% succeeded on C, D, and N. Thus it can be concluded that more Ss in the multiple than single groups were not merely matching ($\chi^2 = 4.20$, $df = 1$, $p < .05$) and further that approximately one-half the Ss succeeded probably by use of rule application.

A further analysis of criterial performance compared relative difficulty among D, DA, and AS trials. Sign tests were used to determine whether one concept compared to another re-

TABLE 4
PERCENTAGES OF ERRORS TO VARIOUS AL-
TERNATIVES IN FOUR TRIALS WITH EACH
CONCEPT

Concept	Alternatives	Single instance	Multiple instance
C	X / Y	40	59
	$\bar{X} \cdot Y$	60	41
D	X · Y	40	55
	$\bar{X} \cdot \bar{Y}$	60	45
N	X · Y	20	12
	X / Y	80	88
AS	Y	10	14
	$\bar{X} \cdot Y^a$	52	52
	$\bar{X} \cdot \bar{Y}^a$	39	34
DA	$\bar{X}^a$	68	64
	X / Y ^a	11	13
	X · Y	21	23

^a An alternative on only two of the four trials.

sulted in more, less, or the same number of errors. Absence of a difference in errors between D and each of the other concepts could be taken as evidence that Ss had learned implicitly during D-trials the exclusive disjunctive rule. Scores from single and multiple-instance groups were evaluated separately; thus, six Sign tests were made. Only one of these resulted in a nonsignificant difference, that being for the comparison between D and DA concepts for multiple-instance Ss ($p > .10$) indicating equal success on D and DA for these Ss.

The final aspect of criterion performance to be considered was an analysis of kinds of errors or misidentified instances made per concept. With C and D concepts, errors were distributed equally between the two alternatives as can be seen in Table 4. For N, AS, and DA, however, disproportionate distributions among response alternatives can be

seen. For N the dominant error was to respond with D but only rarely with C.

With AS the error distributions indicated a disproportionate response to the $\bar{X} \cdot Y$ alternative, which was presented on two trials and accounted for 52% of all errors made in the four trials on which this concept was presented. Recall that these AS instances showed dual elements which in fact instanced more than was necessary for single affirmation. Only 10% and 14% of the errors in single and multiple conditions, respectively, were made with a response which presented only one symbol. These two results taken together indicate that Ss who did not understand affirmation were most likely attempting to match a dual-element instance with a dual-symbol expression.

These data agree with error analyses of DA which showed a majority of errors with dual-element symbol expressions rather than with unitary symbols. In other words, of the DA errors made, the dominant tendency was to match a unitary element with a single symbol even though this symbol was the negation of the element shown in the instance (e.g., *circle* as the instance with $\bar{C}$).

DISCUSSION

This investigation was undertaken to study characteristics of adolescents' concept attainment under two modes of instance presentation. The multiplicity of instances to which Ss were exposed in the attainment series proved to be an important factor determining success in identification of new instances in the criterial series, as multiple-instance exposure led to greater success on four of the five criterial concepts than single-instance exposure. A large proportion of this success might be attributed to the

learning of rules for the logical connectives, but some success seems due to a matching strategy in which Ss compared elements of instances with symbols of logical expressions.

Evidence for rule learning of connectives was inferred from the pattern of success for Ss showing correct identification of C, D, and N, as well as from their equivalent success on D and DA trials. Across all S groups a consistent ordering of concepts from least to most difficult was $C < D = N < DA < AS$. For approximately one-half of the Ss—56% and 35% of the multiple and single instance groups, respectively—this ordering was modified with C, D, and N being equally easy. This pattern of success can be seen as the result of Ss having learned relational rules for the three connectives while it cannot be attributed to the sublogical process of matching. This is to say, with attributes and alphabet symbols defined, an S could have succeeded on C and N trials by matching point-for-point elements of instances with symbols of expressions. With D trials, in distinction, each instance contained one element which had no symbolic equivalent in the expression; likewise, each expression had one symbol for which there was no element present. Since C, D, and N comprise a logical set while they are nonequivalent from the matching standpoint, success with all three operations would seem to be indicative of rule learning.

A second line of evidence for rule learning is seen in performance on DA trials relative to D trials by Ss given multiple-instance training. Recall that DA trials were unique in this experiment in their presentation of instances containing only one element. As with D, a matching strategy would be inadequate for DA. In spite

of this drawback and the novelty of procedure no more errors were made on DA than on D trials by Ss performing with multiple instances. Apparently these Ss had learned the rule for the D connective and were as able to apply it to dual-element instances as to novel unitary instances.

Further evidence for rule learning might have been adduced had success on AS trials not been so minimal. The error analysis for AS trials indicated a strong tendency to attempt a match, even to the extent that a denial of a present element was responded to if the expression containing the denial had two symbols. This preponderance of attempts to match need not be interpreted as incongruous with rule learning, however, since success on DA was probably effected through the rule for the D-connective while in AS trials no connective was given.

These data concur with previous observations on concept attainment reported for college students. Enhanced rule-learning following multiple-instance training agrees with data presented by Turnure and Wallach (1965) who found greater generalization to new instances after varied than after unitary instance training. Our data agree with their interpretation, that the constant critical feature in a set of instances, in our case the connective, is likely to be differentiated from its surroundings more than a stimulus occurring repeatedly in the same context. In our procedure varied instances made explicit the relational features of the connectives while behavior relatively more tied to explicit content was produced with single instancing.

Finally, the results resemble those reported by Haygood and Bourne (1965) for college students. Their Ss

were submitted to a series of instance identification tasks with attributes defined beforehand. As the series progressed, rule-learning as measured by intertask improvement became evident. Haygood and Bourne employed concepts equivalent to C, D, and N of the present study and observed "maximum" rule learning by the second or third problem. Their study varied instances in successive tasks while ours did so in one task, but the performance of our adolescents is similar to their college students. The major difference between results of the two studies would, however, show that not all our adolescents are successful rule learners while from Haygood and Bourne's data (cf. Table 3,

p. 183) it appears that most if not all college students succeeded by the third series.

REFERENCES

- FURTH, H. G., & YOUNISS, J. The influence of language and experience on discovery and use of logical symbols. *British Journal of Psychology*, 1965, **56**, 381-390.
- HAYGOOD, R. C., & BOURNE, L. E., JR. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review*, 1965, **72**, 175-195.
- TURNURE, C., & WALLACH, L. The influence of contextual variation on the differentiation of parts from wholes. *American Journal of Psychology*, 1962, **73**, 481-485.
- YOUNISS, J., & FURTH, H. G. Attainment and transfer of logical connectives in children. *Journal of Educational Psychology*, 1964, **55**, 357-361.

(Received July 28, 1966)

CONTROL OF HYPERACTIVITY BY SOCIAL REINFORCEMENT OF ATTENDING BEHAVIOR

K. EILEEN ALLEN, LYDIA B. HENKE, FLORENCE R. HARRIS,

University of Washington

DONALD M. BAER, AND NANCY J. REYNOLDS

University of Kansas

A 4½-yr.-old boy with excessively short span of attention was helped to acquire more extended attending behavior through the systematic programming of contingencies for adult social reinforcement. When the child remained with a single activity for 1 continuous minute, teachers immediately gave attention and approval for a long as he remained with that activity. Teachers withheld their attention consequent upon all other behavior. Within 7 days the number of activity changes decreased markedly. Reversal of these procedures reinstated the hyperactive behavior. When original reinforcement contingencies were reintroduced, there was again a marked decrease in number of activity changes. The study gives evidence that adults can help a child to increase his attending behavior, a crucial aspect of learning.

There exists now a series of experimental field studies applying reinforcement principles to problem behaviors of preschool children. These studies have dealt with crying (Hart, Allen, Buell, Harris, & Wolf, 1964), regressive crawling (Harris, Johnston, Kelley, & Wolf, 1964), isolate play (Allen, Hart, Buell, Harris, & Wolf, 1964), passivity (Johnston, Kelley, Harris, & Wolf, 1966), non-cooperative behaviors (Hart, Reynolds, Brawley, Harris, & Baer, 1966), self-mutilative scratching (Allen & Harris, 1966), autistic behavior (Brawley, Harris, Peterson, Allen, & Fleming, 1966; Wolf, Risley, & Mees, 1964) and classroom disruptiveness (Allen, Reynolds, Harris, & Baer, 1966). In each instance, the behavior under examination was highly responsive to adult social reinforcement. The present study was conducted to ascertain whether similar social reinforcement procedures could alter the hyperactivity of a 4-year-old boy who tended to flit from activity to activity.

Attending behavior, commonly referred to as "attention span," has long been recognized as a crucial and desirable alternative to hyperactivity. What has not always been clear is the extent to which attending is a behavior which teachers can help a child to develop, although Patterson (Patterson, Jones, Whittier, & Wright, 1965) has done work in this area with older children. Thus it is of interest to determine if systematic social reinforcement can increase the duration of a young child's attending to an activity, and also to analyze the successive steps a teacher might take in helping a child to maintain his attention to an activity for increasingly long periods.

One of the ultimate objectives of preschool education is, of course, to develop a child's skills in using materials constructively and creatively. An essential first step toward this objective sometimes must be to increase the time the child spends engaging in each activity. Fortunately, duration of attention can be defined,

observed, and reliably recorded in the field situation.

METHOD

Subject

James was one of 16 normal children of middle-socioeconomic status who comprised the 4-year-old group in the Laboratory Preschool. At the inception of the study, he was 4 years, 6 months old and had been attending school for 3 months.

James was a vigorous, healthy child with a well-developed repertoire of motor, social, and intellectual skills. Although he made a comfortable adjustment to school during the first few weeks, a tendency to move constantly from one play activity to another, thereby spending little time in any one pursuit, was noted early. Since such behavior is common to some young children in a new situation, his teachers merely continued their friendly efforts to engage him in more prolonged and concentrated use of materials.

After 12 weeks James showed no diminution in number of activity changes during play periods. An observer then was assigned to record his behavior, noting his activities and the time spent in each. Records kept over 5 school mornings showed that although occasionally James stayed with an activity for 1, 2, or 3 minutes, the average duration of an activity was less than 1 minute. The parent reported that the same kind of "flightiness" had long caused concern at home. It was agreed that a study be made of ways of helping James to increase his attending behavior.

Procedure

The procedure for increasing the duration of time spent in any activity was to make adult social reinforcement contingent solely on the subject's (*S*'s) emitting attending behavior for a specified minimum period of time. Attending behavior was defined as engaged in one activity. This included play activity (*a*) with a single type of material, such as blocks or paint; (*b*) in a single location, such as in the sandbox or at a table; or (*c*) in a single dramatic role, such as sailor or fireman. Adult social reinforcement (Bijou & Baer, 1965) was defined as one or more of the following teacher behaviors: talking to *S* while facing him within a distance of 3 feet, or from a greater distance using his name; touching *S*; and giving him additional materials suitable to the ongoing

activity. Withholding or withdrawing social reinforcement consisted of turning away from *S*; not looking or smiling at him; not speaking to him; and directing attention to some other child or activity.

One teacher was assigned major responsibility for maintaining reinforcement contingencies. However, since the two other teachers might at times also deliver or withhold reinforcement, each had to remain constantly aware of the conditions in force.

The design of the study required four successive experimental stages, as delineated by Harris (1964).

Base line. The existing rate, or operant level, of activity changes prior to systematic application of adult social reinforcement was recorded for several play sessions.

Reinforcement. Social reinforcement was presented immediately when attending behavior had been emitted for 1 unbroken minute. Reinforcement was maintained continuously until *S* left the material or the area or verbalized a change in his play role. Immediately consequent upon such a shift in play activity, social reinforcement ceased until 1 minute of attending behavior had again been emitted. The procedure was continued until attending behaviors had materially increased.

Reversal. Then, to ascertain whether social reinforcement was in fact the determining factor in modifying the behavior under study, reinforcement was again delivered on a noncontingent basis such as had been in effect during the base-line period. This reversal of contingencies was carried out long enough to yield a clear assessment of the effects of the changed conditions.

Reinstatement. During this period, the procedures in effect during the second stage, Reinforcement, were reinstituted. After attending behaviors had again increased in duration, the criterion for presenting social reinforcement was raised to 2 minutes.

Recording

The *S*'s attending behavior and adult social reinforcement, as previously defined, were coded and recorded in successive 10-second intervals by an observer using a stopwatch and a red flashlight with a magnet attached. The recording system was similar to that described by Allen et al., (1964). Each period of attending to one activity was enclosed in brackets. Since an increase in attending behavior brought a corresponding decrease in the number of activity changes, data on attending behavior were

counted and graphed in terms of the number of activity changes occurring within successive 50-minute time units. In general, but not necessarily, two 50-minute periods indicated 1 day of recording of play time exclusive of teacher-structured or teacher-directed activities.

During the two reinforcement stages, the observer used a flashlight to inform teachers when *S* reached criterion for social reinforcement. The cue consisted of placing the flashlight on top of the metal clip of the clipboard as soon as *S* had emitted 1 minute, and later 2 minutes, of attending behavior. When the behavior stopped, the observer removed the flashlight and placed it under the clipboard, where it remained out of sight until criterion attending behavior had again been emitted. Teachers were instructed to maintain awareness of the flashlight position and to check it before giving *S* any social reinforcement.

Periodically throughout the study, observer reliability on attending behaviors, activity changes, and adult social reinforcement was checked by an independent observer. Agreement of records ranged between 97% and 100%. No post checks of attending behavior could be made because the study was terminated by the close of the school year, at which time the family moved to another city.

In addition to the behavior under study, some assessment of whether social aspects of *S*'s behavior were affected by changes in his attending behavior seemed desirable. Therefore, *S*'s verbalizations, proximity to, and cooperation with other children were defined, coded, and recorded. The quality of the child's social behavior was estimated by considering cooperative behavior as high-level social behavior and mere proximity as low-level social behavior, in contrast to isolate behavior, which was considered non-social. Interrater reliability on these parameters ranged between 84% and 92%.

RESULTS

Base line—Stage 1. The number of activity changes that James made in each of 21 successive 50-minute periods of free-choice play, both indoors and out, is shown in Figure 1. The fewest number of activity changes were 33 during Period 12, with an average duration of 1 minute 29 seconds per activity. The greatest

number of activity changes occurred in Period 14, with 82 changes and an average duration of 37 seconds per activity. The overall average for the base-line stage (the operant level of the behavior under study) was 56 activity changes per 50-minute period, with an average duration of 53 seconds per activity.

The amount of teacher reinforcement presented to James on a random, noncontingent basis averaged 16% of each session. This rate was within the normal range in this pre-school of amount of teacher attention per child.

Reinforcement—Stage 2. This stage comprised seven 50-minute periods, as shown in Figure 1. Activity changes ranged from a high of 41 in Period 22 (the first period of experimental procedures) to a low of 19 in Period 28 (the last period of experimental procedures). The overall average of activity changes for the seven periods was 27, with an average duration of 1 minute 51 seconds per activity, or twice that of the base-line stage. Teacher reinforcement during Stage 2 averaged 38% of each period.

Reversal—Stage 3. During the four-period reversal stage (Figure 1, Stage 3), activity changes rose markedly. An average of 51 activity changes per period occurred, with an average duration of 59 seconds per activity. Both measures (number of changes and average duration) were comparable to the base-line stage. During the reversal teacher attention averaged 14% of each period.

Reinstatement—Stage 4-A and B. Reinforcement contingencies during Stage 4-A, Figure 1, were the same as those in effect during Stage 2. Under these conditions, the rate of activity changes again dropped markedly, with a high of 31 and a low of 12 (Periods 33 and 36, respectively).

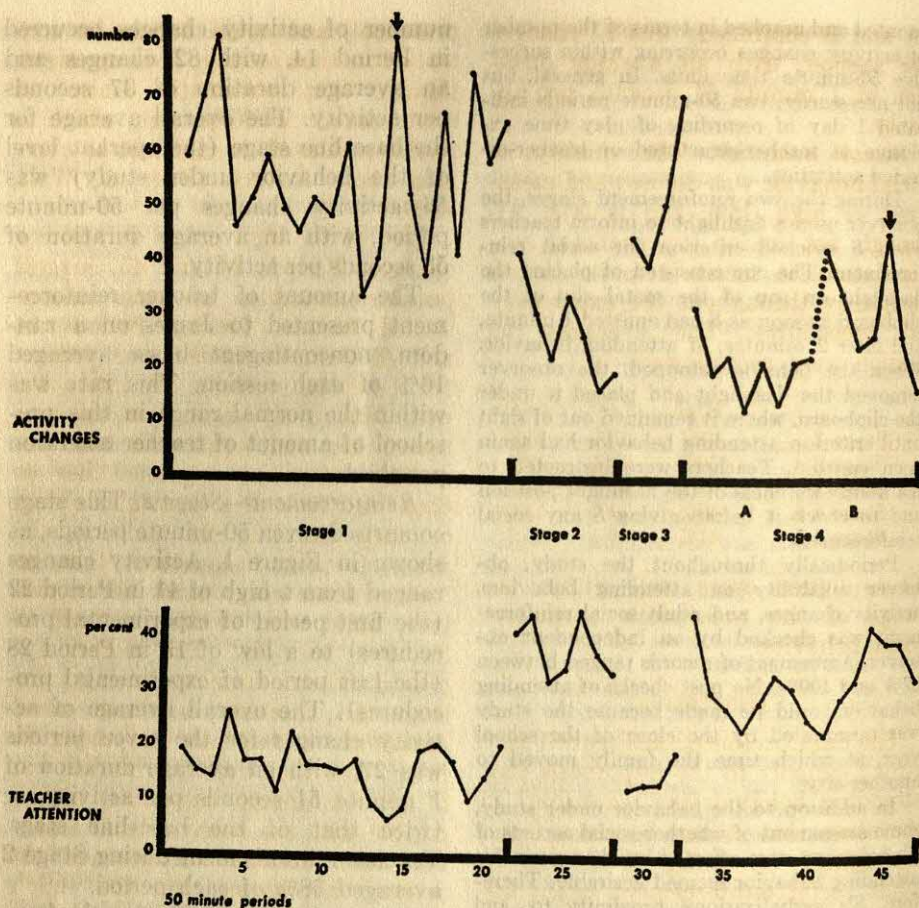


Fig. 1. Number of activity changes of *S* during 50-minute periods throughout study, compared with concurrent percentage of time *S* received teacher attention. (Stage 1, base line of activity change under noncontingent attention. Stage 2, attention contingent on 1 minute of attending. Stage 3, base-line condition. Stage 4, contingent attending as in Stage 2; at dotted line, criterion for attending raised to 2 minutes. Arrows indicate days *S*'s mother visited.)

The overall average of activity changes for the eight periods of Stage 4-A was 20, with an average duration of $2\frac{1}{2}$ minutes per activity. Teacher reinforcement during Stage 4-A averaged 31% of each period.

In Period 41 (Figure 1, Stage 4-B) the criterion for delivery of social reinforcement was raised to 2 minutes of attending behavior. Some increase in number of activity changes occurred during Period 41, with a

subsequent leveling off. During this part of Stage 4, the greatest number of activity changes was 45, occurring in Period 45; the fewest number, 16, occurred in Period 47. The average number of activity changes during Stage 4-B was 32 per period, with an average duration of 1 minute 34 seconds per activity. Teacher reinforcement averaged 33% of each period.

The overall average for Period 4

(A and B combined) was 27 changes per session, with an average duration of 1 minute 51 seconds per activity.

Social behavior. The quality of social behavior was defined and measured as high, low, and isolate. Although not under experimental manipulation, it merits remark for its constancy throughout the experimental procedures. The averages per session were as follows:

	High	Low	Isolate
Stage 1	45%	37%	16%
Stage 2	50%	38%	12%
Stage 3	48%	40%	13%
Stage 4-A	49%	39%	13%
Stage 4-B	41%	44%	15%

These figures were well within the range of the preschool's normative social behavior.

DISCUSSION

The data presented in Figure 1 give strong support to the hypothesis that attending behavior is "teachable" in the sense that it can be shaped and maintained by teachers. Moreover, adult social reinforcement again appears to be a powerful instrument for this purpose. When adult social reinforcement was given in a systematic fashion, solely as an immediate consequence of continuing attending behavior, the number of activity changes diminished to half the number that occurred under the more usual, nonsystematic adult procedures of the base-line and reversal stages.

The continuing fluctuation of the data which occurred during each of the experimental periods may merit comment. Behavior does, of course, vary somewhat from day to day. The factors responsible for this variability were not brought under experimental control. Many of them are inherent in the field setting of a

preschool, and could hardly be controlled in that setting. It is apparent, though, that systematic control of adult social reinforcement, which is readily achieved, is sufficient to override these factors (Baer & Wolf, 1966).

Two of the high points in activity changes, Periods 14 and 45, suggest possible examples of such uncontrolled factors. During Periods 14 and 45 James's mother was present for the entire morning. She interacted with him freely each time he contacted her and went with him frequently when he requested her to come and look at a particular object or play situation. In addition, she made frequent suggestions that he "settle down" and paint her a picture, build with blocks, or "tend to his own business." The mother appeared to have more reinforcing value than the teachers on these novel occasions, a fact not surprising in itself. The fact that the mother was often reinforcing behaviors incompatible with the behavior that teachers were shaping strengthened the original hypothesis that the child's short attention span was in fact a function of adult social reinforcement.

No formal attempt was made to secure data on the quality of James's attending behaviors. In the judgment of the teachers, however, the quality improved steadily. During Stage 4, James frequently spent 15 to 20 minutes pursuing a single activity such as digging, woodworking, or block building. Within these activities he made frequent excursions to get additional materials relevant to his project, such as a wheelbarrow or a dirt sifter. By definition, such departures were recorded as activity changes, even though he returned and continued with the same play. Such occasions, clearly delineated in

the data, teachers considered evidence of improved quality of attending, for the side trips were relevant to a core activity, rather than a series of unrelated activity changes as were typical of Stages 1, 2, and 3. The data thus are probably a conservative estimate of the degree of change produced in James's attention span.

The data on social behavior are of particular interest for they answer in part the often-asked question regarding peripheral effects on overall behavior patterns when one aspect of behavior is under intensive treatment. As was indicated, there was no change in the quality of James's social interaction, already deemed satisfactory by teachers at the start of the study, though the number of separate contacts did decrease, as was predicted. These data add to the evidence that only the behavior specifically being worked on increases or decreases as a function of the reinforcement contingencies.

Throughout the study, the child's mother was informed of procedures and progress in frequent parent conferences. However, no systematic attempts were made to program presentation of social reinforcement from the family. For one thing, the mother worked, and there were frequent changes of babysitters. Nevertheless, the mother reported that James had "settled down" considerably at home. She kept no data to substantiate these statements, but did relate several incidents which indicated that there was some generalization from preschool to home. Both the mother and the teachers judged that James was eminently more ready for kindergarten at the end of the study than he had been prior to it. The importance of intensive attending behavior to future learning is obvious.

The ease of socially altering attending behavior in *either* direction, while perhaps less obvious, is no less important to an analysis of children's intellectual, perceptual, and social development.

REFERENCES

- ALLEN, K. E., HART, B. M., BUELL, J. S., HARRIS, F. R., & WOLF, M. M. Effects of social reinforcement on isolate behavior of a nursery school child. *Child Development*, 1964, **35**, 511-518.
- ALLEN, K. E., & HARRIS, F. R. Elimination of a child's excessive scratching by training the mother in reinforcement procedures. *Behaviour Research and Therapy*, 1966, **4**, 79-84.
- ALLEN, K. E., REYNOLDS, N. J., HARRIS, F. R., & BAER, D. M. Elimination of disruptive classroom behaviors of a pair of preschool boys through systematic control of adult social reinforcement. Unpublished manuscript, University of Washington, 1966.
- BAER, D. M., & WOLF, M. M. The reinforcement contingency in preschool and remedial education. Paper presented at the meeting of the Carnegie Foundation Conference on Preschool Education, Chicago, January 1966.
- BIJOU, S. W., & BAER, D. M. *Child development*. Vol. 2. New York: Appleton-Century-Crofts, 1965.
- BRAWLEY, E. R., HARRIS, F. R., PETERSON, R. F., ALLEN, K. E., & FLEMING, R. E. Behavior modification of an autistic child. Unpublished manuscript, University of Washington, 1966.
- HARRIS, F. R., WOLF, M. M., & BAER, D. M. Effects of adult social reinforcement on child behavior. *Young Children*, 1964, **20**, 8-17.
- HARRIS, F. R., JOHNSTON, M. K., KELLEY, C. S., & WOLF, M. M. Effects of positive social reinforcement on regressed crawling in a preschool child. *Journal of Educational Psychology*, 1964, **55**, 35-41.
- HART, B. M., ALLEN, K. E., BUELL, J. S., HARRIS, F. R., & WOLF, M. M. Effects of social reinforcement on operant crying. *Journal of Experimental Child Psychology*, 1964, **1**, 145-153.
- HART, B. M., REYNOLDS, N. J., BRAWLEY, E. R., HARRIS, F. R., & BAER, D. M. Effects of contingent and non-contingent social reinforcement of the isolate behavior of a nursery school girl. Unpublished manuscript, University of Washington, 1966.

- JOHNSTON, M. K., KELLEY, C. S., HARRIS, F. R., & WOLF, M. M. An application of reinforcement principles to development of motor skills of a young child. *Child Development*, 1966, **37**, 379-387.
- PATTERSON, G. R., JONES, R., WHITTIER, J., & WRIGHT, M. A. A behavior modification technique for the hyperactive child. *Behaviour Research and Therapy*, 1965, **2**, 217-226.

- WOLF, M. M., RISLEY, T., & MEES, H. Application of operant conditioning procedures to the behavior problems of an autistic child. *Behaviour Research and Therapy*, 1964, **1**, 305-312.

(Received August 19, 1966)

SCHOOL-RELATED ATTITUDES OF CULTURALLY DISADVANTAGED ELEMENTARY SCHOOL CHILDREN¹

DANIEL C. NEALE AND JOHN M. PROSHEK

University of Minnesota

A version of the Semantic Differential was used to sample attitudes of 350 children in the 4th, 5th, and 6th grades of 2 elementary schools. Compared to the other schools in the same city, School 1 was low on socioeconomic indicators, School 2 was near the median. Children in School 1 had significantly higher evaluative scores for "my school books," "having to keep quiet," "following rules," and "my school building." Children in School 2 were significantly more positive toward "my teacher," "father," and "college student." As grade in school increased, evaluative scores became significantly less positive for a variety of stimuli, including "my school books," "my classroom," "my teacher," and "me." Attitudes toward several stimuli were similar for boys and girls in School 1 but markedly different for boys and girls in School 2.

A common objective of educational programs for culturally disadvantaged children is the improvement of attitudes toward school and toward learning. Such children presumably have a constellation of attitudes that severely handicaps their school performance, for example, low evaluation of self, low level of aspiration, and negative feelings about school and school work.

The evaluation of educational programs for the disadvantaged is difficult because such attitudinal objectives are difficult to measure. An added problem is the fact that disadvantaged children, almost by definition, have difficulties in responding to existing instruments, which may involve complicated instructions or require good reading and writing skills.

The present study was undertaken

to explore the usefulness of the Semantic Differential in measuring school-related attitudes of culturally disadvantaged elementary school children. The Semantic Differential is extremely flexible and, while it is a paper and pencil instrument, nevertheless, its reading and writing requirements are minimal. Recently, it has been shown to yield stable factor scores with children as low as Grade 2 (DiVesta, 1966; DiVesta & Dick, 1966).

The study was also designed to test some prevailing assumptions about attitudes held by culturally disadvantaged children. In particular, it was hypothesized that children from a school in a neighborhood low on socioeconomic indicators would make less favorable evaluations of a variety of school-related concepts than would children from a school in a neighborhood higher on the same indicators.

METHOD

Two elementary schools in Minneapolis, Minnesota, were chosen as sites for the study. One school was chosen from the local Youth Development Project's "target area";

¹ Research supported in part by grants to the University of Minnesota, Center for Research in Human Learning, from the National Science Foundation (GS541), the National Institute of Child Health and Human Development (1-P01-HD-01136-01-HDP), and the Graduate School of the University of Minnesota.

the other school was selected from a list of "comparison schools," studied in the Project as a contrast to the target area schools.

In the target area about one-third of the residential buildings were rated as dilapidated or deteriorated. Although it contained less than 8% of the city's population, the area had one-third of the families on relief. One-fourth of the families had an annual income under \$3,000. The area has also been characterized by high rates of crime and delinquency, high levels of family mobility, and low rates of owner occupancy. Studies conducted by Project staff members have documented the contrast between target area and comparison school areas on these and other factors (Faunce, Bevis, & Murton, 1965).

In Table 1 are selected socioeconomic indicators for the two schools in the study. Using a composite index which takes into account rental rates, percentage of owner occupancy, school attendance rates, and pupil turnover rates, the first school was at the fifth percentile rank and the other school at the fifty-sixth percentile rank compared to other city schools. A similar contrast between the schools was present in terms of median school years completed, median family income, percentage of non-white students, and a measure of police contacts with youth.

The target area school community proved to be surprisingly high in average income and school years completed. Since the 1960 census, urban renewal projects have forced lower-income families into the district, and therefore the 1960 figures are probably an overestimate. However, one should be cautious about generalizing the results of this study to certain areas of Chicago, New York, or other cities, where the deprivation may be more severe.

The comparison school was selected to represent the average city school. The socioeconomic indicators as reported in Table 1 are consistent with that objective.

A total of 350 children in the fourth, fifth, and sixth grades at the two schools responded on 16 adjective scales to each of 15 stimulus phrases. Following Osgood's procedures (Osgood, Suci, & Tannenbaum, 1957), the responses were subjected to a principal-component factor analysis with varimax rotation to identify clusters of adjective scales that comprised reasonable dimensions of connotative meaning for the subjects in the study.

In Tables 2 and 3 rotated factor matrices

TABLE 1
SELECTED SOCIOECONOMIC INDICATORS FOR
TWO ELEMENTARY SCHOOLS

	Low-SES School	Middle-SES School	City
Composite index (rent, owner occupancy, attendance, and turnover) ^a	1.24 (5 PR)	.94 (56 PR)	1.00
Median school years completed (1960) ^b	9.5	10.5	11.7
Median family income (1960) ^b	\$5,063	\$6,077	\$6,401
Percentage of nonwhite students (1964) ^c	27.5	1.4	8.8
Percentage of 10-17 year olds contacted by police (1964) ^c	11.7	3.7	5.7

^a Report of Minneapolis Public Schools, Department of Administrative Research, April 6, 1962.

^b 1960 Census.

^c Report of Minneapolis Youth Development Project, August, 1965.

are given for the children in the two schools in terms of the three most prominent factors. The evaluative dimension (Factor I) was by far the most prominent, accounting for over one-third of the total variance. Eight adjective scales (noted in Tables 2 and 3) had substantial loadings on Factor I and were selected to constitute the evaluative dimension. Responses of each child on these eight scales were summed for each stimulus phrase to represent the child's attitude toward the stimulus.

A comparison of the factor matrices for the two schools indicates similar loadings of adjective scales on the evaluative dimension. The only exception is the "excited-calm" scale, which loaded more heavily on Factor I in the low-socioeconomic status (SES) school than in the middle-SES school.

The secondary factors (II and III) were not nearly as prominent as the evaluative dimension. Each accounted for less than 10% of the total variance. The activity dimension, frequently found by Osgood, was relatively clear as Factor II, and the commonly found dimension of potency was a clear Factor III.

Because of the weakness of Factors II and III, scores on these dimensions were not subjected to further analysis. The remainder of the study was concerned only with an analysis of scores on the evaluative, or good-bad, dimension.

A three-way analysis of variance was conducted on evaluative scores for each of the 15 stimulus phrases to test the effects of school, sex, and grade.

TABLE 2

ROTATED FACTOR MATRIX FOR 142 FOURTH,
FIFTH, AND SIXTH GRADE STUDENTS IN
A SCHOOL LOW ON SOCIOECONOMIC
INDICATORS

Adjectives	Factor			h^2
	I	II	III	
slow-fast	.327	-.551	-.074	.416
new-old	.059	.544	-.197	.338
wise-foolish	-.733 ^a	.198	-.086	.581
weak-strong	.559 ^a	-.335	-.244	.484
interesting-boring	-.501 ^a	.571	-.047	.579
sad-happy	.680 ^a	-.437	.101	.663
active-passive	-.292	.656	.097	.525
bad-good	.802 ^a	-.182	.098	.686
usual-unusual	-.571	.041	.070	.332
male-female	.066	.054	.667	.453
excited-calm	.506	.517	.061	.528
colorless-colorful	.339	-.526	.072	.398
fair-unfair	-.713 ^a	.248	-.171	.599
soft-hard	-.115	.131	-.735	.671
worried-relaxed	-.693 ^a	-.010	.119	.495
pleasurable-painful	-.683 ^a	.339	-.168	.609
Eigenvalues	4.570	2.471	1.214	8.255

^a Scored as a part of an evaluative dimension.

TABLE 3

ROTATED FACTOR MATRIX FOR 208 FOURTH,
FIFTH, AND SIXTH GRADE STUDENTS IN
A SCHOOL NEAR THE MEDIAN ON
SOCIOECONOMIC INDICATORS

Adjectives	Factor			h^2
	I	II	III	
slow-fast	.390	-.176	.288	.266
new-old	-.209	.566	.037	.365
wise-foolish	-.750 ^a	.036	-.009	.564
weak-strong	.577 ^a	-.141	.401	.513
interesting-boring	-.728 ^a	.305	-.045	.625
sad-happy	.764 ^a	-.229	.077	.642
active-passive	-.480	.404	-.153	.417
bad-good	.804 ^a	-.067	-.022	.651
usual-unusual	-.556	.020	.017	.310
male-female	.059	-.212	-.537	.337
excited-calm	.128	.795	.080	.654
colorless-colorful	.355	-.251	.311	.286
fair-unfair	-.768 ^a	.129	.125	.622
soft-hard	-.212	-.045	.710	.551
worried-relaxed	.640 ^a	.060	-.153	.437
pleasurable-painful	-.693 ^a	.096	.131	.507
Eigenvalues	5.062	1.457	1.229	7.747

^a Scored as part of an evaluative dimension.

RESULTS AND DISCUSSION

Results of the analysis are summarized in Tables 4-7.

In Table 4 the 15 stimulus phrases are ranked from most to least favorable on the basis of the mean scores of all 350 children in the study. Since

responses on each adjective scale could range from one to seven, and the evaluative dimension consisted of the sum of responses to eight of the adjective scales, the scores could vary from 8 (least favorable) to 56 (most favorable). Most were above the neutral score of 32. The overall ranking of the stimulus phrases was quite conventional. Mother and father were evaluated most favorably; fighting and stealing were evaluated least favorably.

Table 4 also gives the mean scores for the children in the two schools and the F -ratio associated with the comparison of each pair of means. Children in the middle-SES school, compared to those in the lower-SES school, made significantly more positive evaluations of "father," "college student," and "my teacher."

Differential evaluation of "father"

TABLE 4

MEAN SCORES ON EVALUATIVE DIMENSION
OF SEMANTIC DIFFERENTIAL FOR CHILDREN
IN TWO SCHOOLS DIFFERING
ON SELECTED SOCIOECONOMIC
INDICATORS

Stimulus	School			Analysis of variance between schools F
	Total	Low-SES	Middle-SES	
Mother	48.0	48.1	48.0	.05
Father	47.7	45.9	49.0	12.10**
College student	45.5	44.2	46.4	6.11*
My teacher	45.5	43.4	46.9	9.72**
Reading a book	45.1	44.1	45.8	2.94
My school building	43.5	45.7	42.0	16.17***
My classroom	42.5	42.7	42.4	.34
Me	41.9	42.0	41.8	.48
My school books	41.6	43.0	40.7	7.07**
Following rules	40.5	41.3	40.0	2.92
Working arithmetic problems	39.7	40.2	39.3	1.51
Talking in front of class	37.9	39.0	37.2	5.64*
Having to keep quiet	32.3	34.6	30.7	15.05***
Fighting with other children	25.7	26.1	25.4	.37
Stealing things	18.0	18.9	17.4	2.69

Note.—For the low-SES school, $N = 142$; for the middle-SES school, $N = 208$; total, $N = 350$. Scores may range from 8 (least favorable) to 56 (most favorable). A score of 32 is neither favorable nor unfavorable.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

may reflect the absence of strong fathers in many low-SES homes. In these data the difference is more pronounced for girls than for boys (see Table 7). Differential evaluations of "college student" and "my teacher" are consistent with the prevailing assumptions that middle-SES children regard college and teachers more favorably than do lower-SES children.

On the other hand, children in the lower SES-school, compared to the middle-SES school, made significantly more positive evaluations of "my school building," "my school books," "talking in front of class," and "having to keep quiet." These findings do not fit prevailing assumptions well.

The relatively more favorable evaluation of "my school building" may reflect the contrast between the school and other buildings in the low-SES neighborhood. Whatever the reason, the marked contrast between

TABLE 5
MEAN SCORES ON EVALUATIVE DIMENSION
OF SEMANTIC DIFFERENTIAL FOR
CHILDREN IN THREE GRADES

Stimulus	Grade			Analysis of variance between grades F
	Fourth	Fifth	Sixth	
Mother	49.0	47.8	47.3	2.05
Father	48.3	47.2	47.6	.26
College student	46.3	45.3	44.9	.86
My teacher	47.8	45.4	43.2	6.98**
Reading a book	45.6	45.2	44.6	1.15
My school building	44.7	43.6	42.2	2.60
My classroom	45.1	43.2	39.4	11.21***
Me	44.0	41.8	40.0	9.05***
My school books	43.4	42.8	39.9	10.06***
Following rules	42.7	40.9	38.1	9.32***
Working arithmetic problems	42.1	40.0	37.1	6.01**
Talking in front of class	41.5	36.6	35.7	9.95***
Having to keep quiet	34.9	32.2	29.8	7.78***
Fighting with other children	27.3	26.5	23.5	3.93*
Stealing things	19.6	18.3	16.2	3.74*

Note.—For the fourth grade, $N = 117$; for the fifth grade, $N = 111$; for the sixth grade, $N = 122$. Scores may range from 8 (least favorable) to 56 (most favorable). A score of 32 is neither favorable nor unfavorable.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 6
MEAN SCORES ON EVALUATIVE DIMENSION
OF SEMANTIC DIFFERENTIAL FOR
BOYS AND GIRLS

Stimulus	Boys	Girls	Analysis of variance between sexes F
Mother	46.2	49.5	16.71***
Father	46.8	48.5	2.25
College student	44.6	46.3	2.72
My teacher	44.2	46.5	5.63*
Reading a book	43.5	46.5	7.64**
My school building	41.5	45.2	11.35**
My classroom	41.3	43.6	4.17*
Me	41.6	42.2	.66
My school books	40.6	42.5	4.50*
Following rules	39.2	41.6	6.28*
Working arithmetic problems	39.6	39.8	.01
Talking in front of class	37.1	38.6	1.69
Having to keep quiet	31.7	32.7	.98
Fighting with other children	27.4	24.3	4.12*
Stealing things	19.3	16.9	2.78

Note.—For boys, $N = 150$; for girls, $N = 191$. Scores may range from 8 (least favorable) to 56 (most favorable). A score of 32 is neither favorable nor unfavorable.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

schools occurs in the responses of boys rather than girls (see Table 7).

The relatively more favorable evaluation by lower-SES children of school books is even more surprising. The fact that "reading a book" is not given a higher evaluation by low-SES children compared to middle-SES children may mean that low-SES children do *value* school books although they do not like to read them.

The fact that "talking in front of class" and "having to keep quiet" are given relatively higher evaluations is likewise puzzling. Perhaps in the lower-SES schools more emphasis must be placed upon the importance of these school virtues, which are perhaps more taken for granted in middle-SES schools.

These findings seem to parallel those of Greenberg, Gerver, Chall, and Davidson (1965), who found a tendency for low-achieving Negroes in a deprived school to be more favor-

TABLE 7
MEAN SCORES ON EVALUATIVE DIMENSION OF SEMANTIC DIFFERENTIAL BY
SCHOOL AND SEX

Stimulus	School				Analysis of variance of School $\times$ Sex <i>F</i>
	Low SES		Middle SES		
	Boys	Girls	Boys	Girls	
Mother	46.7	49.1	45.9	49.8	1.12
Father	45.9	45.8	47.4	50.4	4.37*
College student	43.5	44.7	45.2	47.4	.44
My teacher	42.7	43.9	45.0	48.4	.85
Reading a book	44.0	44.0	42.9	48.3	7.16**
My school building	45.5	45.8	38.9	44.7	11.90***
My classroom	43.1	42.3	40.0	44.5	6.48*
Me	42.4	41.8	41.5	42.5	1.74
My school books	42.5	43.4	39.6	41.8	.65
Following rules	41.1	41.5	38.3	41.8	2.77
Working arithmetic problems	41.9	39.0	38.6	40.4	5.07*
Talking in front of class	39.7	38.5	35.3	38.7	3.64
Having to keep quiet	34.3	34.8	30.0	31.1	.02
Fighting with other children	26.4	25.9	28.7	23.1	3.93*
Stealing things	19.4	18.5	19.7	15.7	2.05

Note.—For low-SES school, $N = 62$ boys, $N = 80$ girls; for middle-SES school, $N = 97$ boys, $N = 111$ girls. Scores may range from 8 (least favorable) to 56 (most favorable). A score of 32 is neither favorable nor unfavorable.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

able than high-achieving Negroes to some socially approved concepts.

Contrary to expectations, children in the two schools did not differ significantly in their evaluations of themselves, fighting, or stealing.

In Table 5 mean evaluative scores are reported for children grouped by grade. A striking pattern is evident in these data, namely, that for every stimulus phrase except one, students in Grade 6 were lower than students in Grade 5 who were lower than students in Grade 4. Only "father" seemed to stem the tide of negativism. Many of the observed trends are highly significant.

The interpretation of this general tendency is difficult. The fact that many of the pronounced differences are related to school, self, and duty may imply a growing dissatisfaction

with school, a kind of progressive alienation. On the other hand, increasingly negative responses as grade in school increases may be a function of shifts in the way children respond to instruments of this type. Further study of this matter is clearly indicated.

In Table 6 mean evaluative scores are given for boys and girls. In every case girls were more favorable than boys to "good" things (mean higher than 32) and less favorable than boys to "bad" things (mean lower than 32). The differences were statistically significant for a number of the school-related stimuli, including "my teacher," "reading a book," "my school building," "my classroom," and "my school books."

Perhaps more interesting than the main effect differences between boys

and girls are the interaction effects between school and sex reported in Table 7. An interesting pattern is evident on a number of the stimuli. Whereas differences between boys and girls are quite marked in the middle-SES school, no such differences exist at the lower-SES school. For example, evaluations of "father" were almost identical for boys and girls in the low-SES school, whereas girls were more favorable to "father" in the middle-SES school. A similar pattern exists for "reading a book," "my school building," and "my classroom." On this instrument, at least, sex differences in attitudes are not as marked for low-SES children as for middle-SES children.

CONCLUSIONS

The interpretation of the data presented in this study is extremely difficult. With such an instrument the subject's responses are to some degree dependent upon his interpretation of what responses are valued by the experimenter. Before much confidence can be placed on interpretations of these data, the instrument should be tested in a variety of contexts to estimate the extent of such bias. However, in spite of such reservations, the results of the present study do provide a number of challenging hypotheses which might well be pursued further.

In the first place comparisons between the two schools indicate that culturally deprived children, as defined in this study, are not negative about school, at least in the sense of devaluing school and school-related activities. To the contrary, it appears that school is valued highly, perhaps as something difficult to attain and perhaps as a place where unpleasant things occur, but nevertheless valued. It is extremely im-

portant to keep in mind that the evaluative dimension on the Semantic Differential is subject to ambiguities on this point. A positive evaluation may reflect high valuation or a strong liking, and the two things do not always go together. A person may easily place a high value on scholarly endeavors but not like to engage in them.

Second, the data in this study suggest a systematic change in attitudes toward school as a function of grade level. Evaluations of a variety of school-related phrases were increasingly negative as grade in school increased. The same was true for an evaluation of self. The trends were not a function of school or sex. One hypothesis that must be considered is that children learn to like school less as they progress through the grades. If so, it suggests the schools are falling down on one of the most important proximate objectives of education, the cultivation of a positive attitude toward schooling. Such an attitude has long been held to be a part of a sound motivation for school learning. In addition, the fact that evaluations of "me" were increasingly negative as a function of grade level implies that children like themselves less as they grow older. A positive self-concept is also held by educators as a desirable objective of schooling.

Of course, these data are far from conclusive. For example, the fact that evaluations of fighting and stealing also became increasingly negative as a function of grade level is not consistent with the view that children are becoming alienated from school and society. Instead it suggests a change in response set to the instrument, as if with increasing age one reports all evaluations somewhat more negatively. In that case the

data do not suggest a failure of schools but rather an important fact to be taken into account in interpreting attitude questionnaire data from different age groups.

Finally, the data in the study contradict the expectation that sex differences in evaluations of school that are common in middle-class children are necessarily found in culturally disadvantaged children. The hypothesis is rather suggested that sex-role differences with respect to education may not be as important in culturally disadvantaged as in middle-class schools.

REFERENCES

DiVESTA, F. J. A developmental study of the semantic structures of children. *Jour-*

nal of Verbal Learning and Verbal Behavior, 1966, **5**, 249-259.

DiVESTA, F. J., & DICK, W. The test-retest reliability of children's ratings on the semantic differential. *Educational and Psychological Measurement*, 1966, **26**, 605-616.

FAUNCE, R. W., BEVIS, D. D., & MURTON, B. *Student mobility in selected Minneapolis schools: Report No. 1*. Minneapolis: Youth Development Project of the Community Health and Welfare Council, 1965.

GREENBERG, J., GERVER, J., CHALL, J., & DAVIDSON, H. Attitudes of children from a deprived environment toward achievement-related concepts. *Journal of Educational Research*, 1965, **59**, 57-61.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.

(Received August 23, 1966)

ACQUISITION OF PROBLEM-SOLVING STRATEGIES IN YOUNG CHILDREN AND ITS RELATION TO VERBALIZATION¹

CAROLYN STERN

University of California, Los Angeles

That young children could learn problem-solving strategies was demonstrated ($p < .01$) with 107 3rd-grade children. A $2 \times 2 \times 2$ factorial design with 2 levels of MA was used. The treatment conditions included multiple-hypotheses testing with and without speaking, and single-hypothesis testing with and without speaking. The problems required the selection of 1 of 4 concepts as the basis for matching an exemplar to a model in each 6-slide problem. Children taught to use knowledge of results to test 1 concept at a time (SH) scored significantly higher on posttests than did those taught to test several hypotheses at once (MH). Verbalization demonstrated no effect with either treatment group; mental age was an important variable only in the acquisition of the MH strategy.

In a series of concept-identification studies reported by Wittrock (1964) and Stern (1965), the authors repeatedly observed that young children tended to perform at purely chance levels when the basis for solution was not provided by the experimenter. Under such "discovery" treatments, children frequently ceased attending to the specific stimuli of the problem situation and either curtailed their efforts or engaged in various types of random activity.

These children did not generate techniques for problem-solving even when the number of trials to criterion was unlimited. On the contrary, since the correct answer was unrelated to a discoverable sequence, for example, some type of alternation pattern, and since the proportion of successes to failures was purely a function of chance, many children gave up in frustration before arriving at a solution. Obviously, the invigorating and intellectually

stimulating effects of discovery, defined as an instructional procedure, were not experienced by those whose best efforts did not result in discovery, defined in terms of performance. Even those children who were able to derive their own problem solutions during training did not subsequently demonstrate significantly improved ability to cope with a related but slightly different type of problem.

There are several ways in which these findings can be interpreted. First, they may be viewed as supporting Piaget's theoretical postulates of sequential cognitive development. From this position it might be argued that young children have not yet reached the stage where they are able to perform the logical operations essential to inferential reasoning. On the other hand, accepting Jerome Bruner's dictum that "any subject can be taught to any child in some honest form" together with the programmer's hypothesis that "there are no poor learners, only poor programs," it might be inferred that a simpler program, with smaller steps and more repetition, would be able

¹This research was supported by the Cooperative Research Program, United States Office of Education (Project No. 2477).

to lead children of this age to discover how to solve concept-identification problems without direct assistance.

Still another way of looking at these findings is that the child's ability to cope with and profit from discovery, as a heuristic, is a function of his previous acquisition of skills of "learning-to-learn." That is, the ability to solve problems can be viewed as an instructional objective, analogous to motor skills or subject matter content. As Gagné (1966) points out, "The strategies of discovery...do not have to be *learned* by discovery [p. 148]."

A series of experiments was designed to test the hypothesis that children can be taught strategies which will subsequently improve their ability to "discover." The first experiment (Stern & Keislar, 1966) used four groups of third-grade children. Two of these groups were given instruction, each in a different problem-solving strategy. The third group was given practice with the materials and problems; presumably members of this group would have the opportunity to derive their own procedures for solving these problems. A fourth group was given preliminary familiarization with the experimental situation but no opportunity to practice with the problems.

The results demonstrated that children taught strategies for solving problems performed significantly better ($p < .001$) with new but similar problems than children who had not been given instruction. This superiority was found only with the simpler of the two instructed strategies. The more sophisticated strategy, which should have produced superior problem-solving skill, was in actuality superior neither to the practice-without-instruction group nor to the no-practice, no-instruction group.

Closer inspection of the data revealed that the performance of these groups was approximately chance; that is, the children had not learned to use the strategy and hence were no different from children who had not been exposed to this training. One source of difficulty might be that the children were unable to cope with the problems of memory storage and information retrieval. Perhaps the children were unable to remember which concepts had been tested and affirmed and which ones rejected. Thus they might be wastefully retesting already disproved hypotheses. A new experiment was therefore designed to test whether young children could be taught to execute this superior strategy if a more effective instructional program, one which also provided some memory storage facilitation, was devised.

Before beginning this experiment, several weeks were devoted to program development. Each phase of the instruction was tried out and tested with about 40 children selected at random from the school population to be used in the experiment. When the final program was set, several new features had been added. The most important of these was a booklet system by which the factor of memory load could be minimized as an effective variable. In addition, the question of the effect of verbalization was raised. In the earlier experiment, all the children had been required to verbalize overtly the correct concept before making a selection response based on that concept. The implied assumption, that verbalization would facilitate learning, was not tested in that design. It was quite possible that verbalization, contrary to expectation, was producing interference, rather than facilitation.

The present experiment was de-

signed to test the following hypotheses:

1. Children taught to follow instructions with a theoretically superior type of problem-solving strategy will be better problem-solvers when these instructions are used as prompts, compared with children prompted with a set of simpler strategy instructions;

2. When strategy prompts are removed, the children taught the more sophisticated strategy will continue to solve the same type of problem more successfully than children taught the simpler strategy;

3. Children who are required to say the correct concept aloud, under either strategy treatment, will be more competent problem-solvers than children who are not taught to supply any overt verbal cues.

METHOD

Apparatus

The UCLA Group Teaching Equipment, described in detail in Stern (1965) with the voice-relay modification reported in Keislar, Stern, and Mace (1966), was used for this study. The basic components included 10 three-sided booths, each equipped with earphones, a microphone, and a multiple-choice individual response panel, all connected to a master control set. A Kodak Carousel slide projector and a Wollensak stereo tape recorder insured the controlled presentation of the audio-visual program; a Clary data recorder, with Flexowriter tape-to-card converter, automatically recorded and interpreted the individual responses.

Subjects

All the third-grade children in a West Los Angeles elementary school (with the exception of approximately 40 children who had participated in the program-development phase) were included in the study. There were 107 children (56 boys and 51 girls) assigned to four experimental treatments, using a stratified-random design based on mental age. The mean chronological age of the experimental subjects was 8 years 5 months, the standard deviation

4 months; the mean mental age was 9 years 4 months, standard deviation 1 year; the mean IQ was 110.6, *SD* 13.0.

Procedure

All instruction and testing were carried out entirely by means of autoinstructional programs, using the equipment and materials described. Before the instruction began, all the children were given a pretest to see how well they could perform the problem-solving task. Then followed 6 days of training, each session lasting approximately 15 minutes. Immediately after the training, a posttest was given; 7 weeks later, a retention test was administered.

Experimental task and materials. The basic task for this study was identical to that described in Stern and Keislar (1967). A 35-mm. color slide presented a model picture in the top center and two exemplars below. A problem consisted of a set of slides, for all of which the basis for matching the exemplar with a model was one of four concepts: number, color, size, or shape. The object was for the child to use knowledge of results to identify which of these four concepts was the basis for matching for each problem.

Experimental treatments. There were two strategy programs, Multiple Hypothesis testing and Single Hypothesis testing, each presented under two conditions of verbalization, not-speaking and speaking, thus providing four experimental treatments.

Within each of the strategy treatments, the only difference between the speaking and not-speaking condition was that the speaking groups were required to say the selected hypothesis aloud, in addition to writing the correct rule. For instance, where the not-speaking groups were told: "Choose a rule and write it in the proper box. Press the button that goes with this picture," the speaking groups were told: "Choose a rule and write it in the proper box. Say the rule. Press the button that goes with this picture." For the speaking condition, the voice-relay mechanism was adjusted so that if the child did not speak he could not receive a reinforcing light for his selection response.

Memory storage. In the previous experiment, it had been observed that children had difficulty retaining negative feedback information so as to consistently reject incorrect solutions. To avoid this problem, children were supplied with booklets and pencils to use in aiding memory.

In the Multiple Hypothesis treatment,

the subjects were helped to focus on the two hypotheses associated with the correct exemplar on the first slide, and then to select one of these two on the basis of information provided with the second slide. Having thus attained the rule by the second slide, they should be able to select the correct exemplar for the remaining slides of the problem. For the Single Hypothesis group, the booklets helped remind each child which concepts were possible and which ones had been tried and rejected, so that a non-replacement procedure could be adopted.

The booklets consisted of four different types of pages appropriate for the needs of each strategy group and each lesson. Figure 1 presents samples of the four page types, reduced in size. The actual booklets were constructed by cutting regular manuscript paper into thirds, width-wise, so that each sheet was approximately $3\frac{1}{2} \times 8\frac{1}{2}$ inches. The number and types of pages varied for each day's lesson, for each of the basic strategy treatments. Table 1 shows the composition of the booklets for both programs.

Strategy instruction. On the first day of training, all treatments received the same instructions and slides for the initial 25 items of the lesson. The first six slides taught the children how to operate the response panels; the next 10 slides presented the concept-identification task in terms of two concepts, color and shape. The booklets were then distributed and the children were taught to write the first letter of the correct concept, as illustrated in nine slides, in the appropriate box on the Type A booklet page (see Figure 1).

The Multiple Hypothesis groups were then shown six slides in which one of the exemplars was the same as the model in both color and shape. As the first basic step in the Multiple Hypothesis testing procedure, the child had to write both of these

TABLE 1
COMPOSITION OF BOOKLETS OVER
SIX LESSONS

Lesson	Multiple hypothesis program	Single hypothesis program
1	5 Pages of Type A 2 Pages of Type B	10 Pages of Type A
2	4 Pages of Type A 2 Pages of Type B	6 Pages of Type C 4 Pages of Type A
3	Same as Lesson 2	4 Pages of Type A 8 Pages of Type D 5 Pages of Type C
4	2 Pages of Type A 4 Pages of Type D 3 Pages of Type B	2 Pages of Type A 5 Pages of Type C 9 Pages of Type D
5		
6	2 Pages of Type B	5 Pages of Type D

rules in the same box, corresponding to the position of the exemplar on the screen. The children were then taught the second basic step in this strategy: to select the one correct concept from the two identified as associated with the correct exemplar. They were given six four-slide problems in which they used the Type B pages.

The Single Hypothesis groups, on the other hand, completed the first lesson with practice in selecting a concept, associating it with a specific exemplar, and testing to see if it was correct, using only Type A pages.

In the first lesson, only two concepts were used, color and shape. The third concept, size, was introduced on the second day and the fourth concept, number, on the third day. After the fourth concept had been identified, a card (Figure 2) was hung on the front panel of each booth, listing the repertoire of rules from which problem solutions could be drawn. This was another way in which the memory load of the experimental task was reduced for all groups.

While the visual stimuli were not identical for the two programs, as they were in the first experiment, the same slides and similar tasks were provided wherever possible. On the fifth and sixth days of training, the slides and problems were identical for all treatment groups; only the strategy instructions varied.

Criterion Tests

The posttest, given after the last day of training, consisted of nine six-slide problems, similar to those used during training. Two months later, the retention test, using the same slides but in different problem sets, was administered.

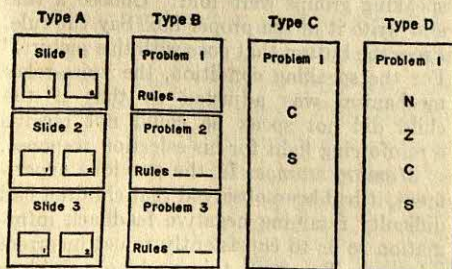


FIG. 1. Sample sheets from booklets for training program.

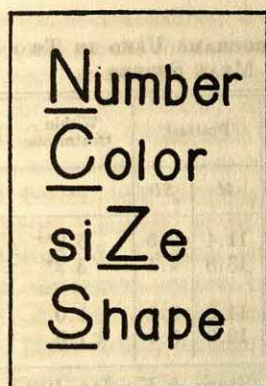


FIG. 2. Rule card.

RESULTS

The means and standard deviations on all measures for various treatment subgroups are presented in Table 2.

Before any strategy training was instituted, all the children were given a problem-solving pretest to determine how well they could perform the experimental task before training. The scores on this two-choice task were at a chance level, since the groups averaged approximately 10 points out of 20. Evidently without training of some kind the problem-

solving task adopted for this project was too difficult for these children.

For each of the experimental groups, the importance of the gain made during training can be assessed in terms of the difference between the scores on the problem-solving pretest and the training test. Here both of the within-treatment differences were significant ($t = 8.4$ and 9.3 , $p < .001$). These gains remained at a reliable level ($t = 4.5$, $p < .01$ and $t = 9.4$, $p < .001$) for the Multiple and Single Hypothesis treatments, respectively, on the posttest which did not provide the procedural steps for each of the strategies.

In order to compare the effectiveness of the new programs with those used in the previous study, the data of that study were reanalyzed and scores on training and posttest obtained using the four test items on each of the last five problems. Table 3 includes these recomputed scores, together with comparable scores from the present study. These data show a significant difference ($t = 3.3$, $p < .01$) in favor of the Multiple Hypothesis treatment on the training test in

TABLE 2
MEANS AND STANDARD DEVIATIONS ON DEPENDENT AND INDEPENDENT VARIABLES

Treatment	Mental age (yrs.-mos.)			Pretest ^a		Training ^b		Posttest		Retention test	
	N	M	SD	M	SD	M	SD	M	SD	M	SD
Multiple hypothesis											
Speaking	27	9-0	1-0	9.7	2.8	17.0	3.0	13.8	4.5	14.7	4.1
Not speaking	28	9-6	1-0	11.7	2.9	15.4	4.1	13.8	3.8	15.3	3.3
Total	55	9-2	1-0	10.8	3.0	16.2	3.7	13.8	4.1	15.0	3.7
Single hypothesis											
Speaking	26	9-4	1-0	11.2	3.3	15.9	2.7	15.6	3.1	15.6	2.2
Not speaking	26	9-6	1-4	9.6	2.4	16.0	3.0	16.5	3.2	16.5	3.0
Total	52	9-5	1-2	10.4	3.0	15.9	2.9	16.0	3.1	16.0	2.7

Note—All tests consisted of 20 items.

^a This test, administered before training, was another form of the posttest.

^b This test was given on the last day of training before the strategy prompts were removed.

TABLE 3

COMPARISON OF THE EFFECTIVENESS OF STRATEGY PROGRAMS USED IN TWO STUDIES,
BASED ON TRAINING AND POSTTEST MEAN SCORES

Treatment	Study ^a	N	Training		Posttest		Within treatments	Between treatments
			M	SD	M	SD	t	t
Multiple hypothesis	1	26	15.7	2.7	11.4	2.5	5.9**	3.3*
Multiple hypothesis	2	55	16.2	3.7	13.8	4.1	3.2*	1.0
Single hypothesis	1	29	13.2	2.9	14.0	3.1	1.0	3.6**
Single hypothesis	2	53	15.9	2.9	16.0	3.1	<1.0	3.2*

^a Study 1 refers to the previously-cited experiment (Stern & Keislar, 1966). Study 2 refers to the present experiment.

* $p < .01$.

** $p < .001$.

the first study, which was not found with the revised programs used in the second study. Comparing the training test and the posttest, however, there was a significant difference between these two tests with the Multiple Hypothesis treatment in both studies. The Single Hypothesis groups showed reliable decrements when their prompts were removed. Evidently the Single Hypothesis groups internalized the instructions they had been given and were able to apply the procedure with or without prompts.

Analyses of variance and covariance were computed, using the pretest score as the covariable in the latter analyses. No differences in significance levels were found, reflecting the fact that the correlations between pretest and posttest scores were .08 and -.10 for the Multiple and Single Hypothesis strategies, respectively. Since the analyses of variance are considered more conservative statistical tests, only these are reported. Significant main effects were obtained for treatment ($p < .01$) confirming the superiority of the Single Hypothesis strategy. No significant differences were obtained for the verbalization variable. The group

which spoke aloud the basis for problem solution was not measurably superior to the group which was not required to verbalize.

Mental age proved to be an important variable in the training and posttest measures. This reflects the fact that the posttest scores for the high MA Multiple and Single Hypothesis treatment groups were 15.0, SD 3.5, and 17.0, SD 2.6, respectively; the low MA groups scored 12.9, SD 4.3 and 15.2, SD 3.4 on these same tests. However, there were no reliable interaction effects with verbalization or treatment which could be attributed to differences in mental age. No differences held up after a 2-month retention period.

While the Single Hypothesis group scored approximately the same on the retention test as they had on the posttest, the Multiple Hypothesis treatment showed an appreciable increment over their posttest scores, achieving a mean score only slightly below that of the Single Hypothesis group.

DISCUSSION

Although this experiment did not include a control group, comparing the score on the pretest for problem-

solving with the scores on the training and posttests demonstrates that children can be taught strategies for solving certain problems, and that they will then be better able to solve problems which they were not able to solve before this training. These data support findings of the earlier study (Stern & Keislar, 1967) in which an untrained control had been included in the experimental design. In that study, neither of the treatment groups had performed as well as could be expected on the basis of mathematical probability. Reexamination of the two procedures revealed that very different tasks were involved. Since the primary objective of the training programs was to bring all the children up to an acceptable level of competence in the use of the treatment strategy before testing for effectiveness of problem-solving with that strategy, it could be argued that it was not necessary to adhere to training programs in which the items were strictly comparable. It was therefore decided to prepare the most effective programs possible for each of the strategy procedures.

The revised program for the Multiple Hypothesis strategy did produce performance comparable to that of the new Single Hypothesis program under the training conditions. However, on the posttest, when the prompts were removed, the Single Hypothesis procedure was again significantly superior to that of the Multiple Hypothesis procedure. Evidently, young children are unable to profit fully from the advantages of the sophisticated strategy, whereas they can use the simpler strategy with optimum efficiency. Thus the first hypothesis was neither supported nor disconfirmed, although there seemed to be some basis to conclude that successively testing hypotheses

produces better problem solving with the population tested than the more economical scanning procedure.

The third hypothesis of the present study was not supported. This is contrary to the findings of Weir and Stevenson (1959), as well as a number of other studies cited in that report, which demonstrated that verbalization concerning the stimuli aided learning at all age levels. These investigators also reported that they had found no evidence to support the prediction that verbalization would be of less value to older children, who, presumably, could use their own implicit verbalizations of the appropriate labels as mediators in solving problems. It is interesting to note that the "older" children referred to were 9 years of age, or approximately the same age (both chronologically and mentally) as the children used in the present study.

As previously noted, all the children in the first study were required to verbalize the problem solution. The fact that the mean scores obtained in that study were generally lower than those of the present study, either with or without verbalization, lends some inferential support to the conclusion that requiring the production of an overt verbal response does not facilitate learning. Perhaps the concepts which were the basis of problem solution were so familiar to all the children that the self-cueing effect of overt speech was of minimal importance.

In conclusion, the present study provides further evidence that young children can be taught strategies which will improve their ability to solve certain types of problems. It seems that a simple strategy is more effective with this age group, and that, with the familiar materials used, requiring the overt verbaliza-

tion of the basis of problem solution provides no reliable difference in the efficiency with which such problem solutions are attained.

The content used for the concept-identification problems was of a simple, primary stimulus nature. The rationale for this selection was to reduce as much as possible any confounding effects of individual differences in degree of familiarity with the material. In this way it would be easier to look at the children's ability to utilize the strategies. An underlying motif of the study was to determine whether children could be taught the logical operations which comprised the component elements of the strategies at younger ages than would be expected on the basis of the Piaget formulations. It also attempted to devise a system of externalizing the children's responses so as to make it possible for the experimenter to engage in formative evaluation, that is, to assess progress at each of the steps in the process.

Whether the strategies taught would have transfer value in quite different types of contexts was not tested. It seems reasonable to assume

that opportunity to practice the strategies with a wide variety of materials would be essential if a generalized application of a problem-solving heuristic is to be attained.

REFERENCES

- GAGNÉ, R. M. Varieties of learning and the concept of discovery. In L. Shulman & E. R. Keislar (Eds.), *Learning by discovery: A critical appraisal*. Rand McNally, Chicago, 1966. P. 148.
- KEISLAR, E. R., STERN, C., & MACE, L. L. Sequence of listening and speaking in beginning French. *American Educational Research Journal*, 1966, **3**, 169-178.
- STERN, C. Labeling and variety in concept identification with young children. *Journal of Educational Psychology*, 1965, **56**, 235-240.
- STERN, C., & KEISLAR, E. R. Acquisition of problem solving strategies by young children and its relation to mental age. *American Educational Research Journal*, 1967, **4**, 1-12.
- WEIR, M. W., & STEVENSON, H. W. The effect of verbalization in children's learning as a function of chronological age. *Child Development*, 1959, **30**, 143-149.
- WITTRICK, M. C. Transfer through verbal cueing in concept identification. United States Office of Education Cooperative Research Project No. 1648, 1964, University of California, Los Angeles.

(Received August 29, 1966)

ACQUISITION AND RETENTION OF DISCRIMINATION LEARNING SETS IN LOWER-CLASS PRESCHOOL CHILDREN¹

PHYLLIS A. KATZ
New York University

Experiment I investigated the effects of programming sequences on the learning set acquisition of lower-class nursery school children. 1 group of 12 Ss was given simple discrimination problems which progressively became more difficult. A 2nd group of 12 Ss received simple problems which abruptly shifted in difficulty level. A 3rd group of 12 Ss received an equivalent number of only difficult problems. The 3 groups did not differ in their performance on criterion problems, however all Ss exhibited positive transfer relative to a control group which did not receive prior training. A significant Treatment $\times$ Day $\times$ IQ interaction indicated that the various programming sequences elicited differential acquisition curves among the normal and low IQ children. The problem sequence appeared to affect the performance of low IQ Ss more than those Ss of normal intelligence. The second experiment assessed the retention effects of the learning set acquisition. A group of 24 Ss was tested 6 mos. after participation in Experiment I and their performance was compared to a control group matched for IQ. The group that had received prior training exhibited faster learning than the control, thus demonstrating retention after a 6-mo. interval.

The phenomenon of discrimination learning sets has recently received a great deal of attention, and an expanding body of literature has demonstrated its occurrence across a wide variety of species, including rats (Koronakos & Arnold, 1957), cats (Warren & Baron, 1956), monkeys (Harlow, 1949; Miles, 1957), marmosets (Miles & Meyer, 1956), chimpanzees (Hayes, Thompson, & Hayes, 1953), and normal and retarded children (Ellis, Girardeau, & Pryer, 1962; Koch & Meyer, 1959; Shepard, 1957). Learning set refers to positive intra-problem transfer which occurs when subjects (Ss) are given practice on an extensive series of similar types of

problems. Research in this area to date has been largely concerned with the role of such gross organismic variables as phylogenetic differences (e.g., Ellis et al, 1962), intelligence level (Jensen, 1963; Kaufman & Peterson, 1958; Koch & Meyer, 1959) and developmental status (Harlow, Harlow, Rueping, & Mason, 1960). Despite the clear relevance of learning sets to educational practice, little experimental attention has thus far been devoted to the issue of how certain manipulable characteristics of the task itself may influence either the speed of learning set formation or its retention. Such issues as how the difficulty of the discrimination problem affects rate of learning set acquisition, what the most efficacious methods of stimulus presentation and programming are, and the kinds of testing conditions which give rise to long-range retention all appear to be in need of investigation.

The major purpose of the present

¹This research was supported by Grant #40-013-1, United States Office of Economic Opportunity, Martin Deutsch, Principal Investigator. Thanks are extended to Erhard Eisner and Ruth Markowitz for their assistance in data collection, and to Public Schools #68, 90, and 175 in New York City for their cooperation in providing subjects.

study was to obtain information regarding the role of task difficulty in the formation of discrimination learning sets by nursery school children. The difficulty level of the discrimination tasks was operationally defined in the present investigation as a function of the number of stimulus attributes common to the two discriminanda. Thus, two stimuli grossly differing with respect to color, shape, and size were regarded as constituting a simple discrimination, whereas two stimuli differing in only one of these attributes were regarded as a more difficult discrimination. This formation is generally in accordance with the theory postulated by Bourne and Restle (1959). If the criterion task consists of difficult discrimination problems, the question may be raised as to which type of training sequence elicits maximal transfer. There are at least three types of programming sequences which might logically be employed. One technique might begin with relatively simple problems which become progressively more difficult. An alternative method is to begin with simple problems and abruptly shift to difficult problems. A third possibility is to begin with difficult problems and maintain the same level of difficulty throughout the training. The present study attempted to assess the effects of these three types of training sequences upon the rate of learning set formation.

If the learning set phenomenon is as important in the educational process as many investigators have noted (e.g., Bruner, 1964; Harlow, 1949, 1959), then it would be expected that learning sets, once formed, should be retained by the organism over relatively long time periods. There has been, however, no experimental demonstration which supports this expectation. Thus, an additional purpose

of the present investigation was to obtain information regarding the retention aspects of learning sets. Towards this end, preschool children were retested 6 months after their initial training in order to assess the degree of retention.

METHOD

Experiment I

The Ss were 48 children enrolled in an experimental nursery school program specifically designed for a culturally deprived population. The mean chronological age of the sample was 52 months. The children were from lower socioeconomic backgrounds, and most of them were Negro. Their Stanford-Binet IQ's ranged from 69 to 111 and the mean IQ was 92.

The stimuli employed consisted of pairs of pictures of geometric figures differing in shape, size, and color. The apparatus employed to present the stimuli has been previously described in detail by Kendler (1959). It consisted of two apertures which exposed the stimuli, and two levers under the apertures. The child was instructed to "press the stick under the picture you think is a winner." If a correct choice was made, a marble was automatically delivered to S in a cup under the stick. For an incorrect choice, S had to return a marble to the experimenter (E). He was given six marbles at the beginning of each problem which were placed in a marble board. The S was instructed that if he won "enough marbles," they could be traded in at the end for prizes, which included modeling clay, water pistols, crayons, jacks, and toy trains. The prizes were exhibited to the children at the beginning of the experimental session.

The Ss were dichotomized with respect to IQ. The low group had a mean IQ of 84.2 whereas the corresponding figure for the high group was 102.4. Within each intelligence classification, Ss were randomly assigned to four experimental treatment groups, which differed according to the kind and number of problems presented during training. For all groups, three criterion problems were administered. These criterion problems, the "difficult" ones, consisted of two discriminanda differing along a single dimension, either color or shape. The various experimental groups can be described as follows:

1. Group I received six easy problems (differing along three dimensions) on Day 1, six problems of intermediate difficulty (differing along two dimensions) on Day 2, and three difficult problems (differing along one dimension) on Day 3;

2. Group II received six easy problems on Day 1, six difficult problems on Day 2, and three difficult problems on Day 3;

3. Group III received six difficult problems on Day 1, six difficult problems on Day 2, and three difficult problems on Day 3;

4. Group IV received only the three difficult criterion problems on one day of testing.

A maximum of 20 trials of each problem was administered. The position of the reinforced stimulus varied between the right and left apertures. This variation was randomly determined in advance with the limitation that the reinforced stimulus did not recur in the same position more than twice in succession. The problem was discontinued after either 20 trials or three consecutive correct choices.

Experiment II

Six months after the original training, 24 Ss were randomly selected from the low and high IQ conditions of Groups I, II, and III in the first experiment. These children comprised the experimental group. A control group of 24 Ss, matched for IQ, was selected from the same nursery schools. These children had not participated in the earlier study.

All Ss participating in Experiment II received 20 trials of three difficult discrimination problems, administered on the same day. The apparatus, instructions, and reinforcement procedure used were the same as described in Experiment I.

RESULTS

Experiment I

The mean number of trials to criterion (i.e., to three consecutive correct responses) for each group on the three criterion problems is presented in Table 1. A comparison of the training groups (Group I, II, and III) with the control group (Group IV) was significant at the .01 level ($F_{1,46} = 7.85$), indicating that the experimental groups were superior on the criterion

TABLE 1
MEAN NUMBER OF TRIALS TO CRITERION
ON CRITERION TASK

Group			
I	II	III	IV
4.32	3.10	3.62	7.32

problems to the group receiving no prior training. The difference among the means of the three training groups, compared by means of an analysis of variance, was not statistically significant. Thus, it may be concluded that all three training groups exhibited positive transfer relative to the control group, however the particular treatment employed did not influence the degree of learning exhibited on the criterion problems.

In order to assess whether or not the particular programming sequences had any effect upon the rate of learning set acquisition, a repeated measures analysis of variance (Lindquist, 1956) was conducted on the mean trials to criterion scores on the three days of testing for Groups I, II, and III. These means are presented in Figure 1. The analysis of these scores revealed that the main effect of day was significant ($F_{2,60} = 9.50, p < .01$). This finding indicates the general improvement exhibited by the total sample on each day of testing. The means for Days 1, 2, and 3 were 6.2, 3.9, and 2.6, respectively. It can be seen in Figure 1 that most of the positive transfer was observable on the second day of testing. The analysis also indicated that the interaction of Day $\times$ Treatment $\times$ IQ was significant ($F_{4,60} = 3.71, p < .05$). This interaction indicates that the various programming sequences elicited differential acquisition rates among the high and low IQ children. As can be seen in

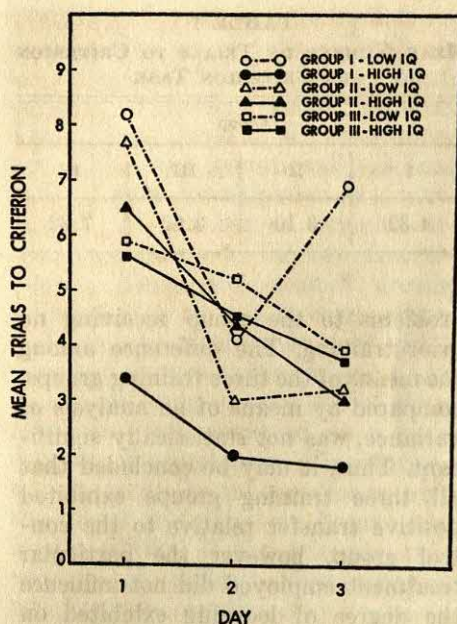


FIG. 1. The mean number of trials required to reach criterion for each group on each day of testing.

Figure 1, the high IQ groups generally show a gradual improvement over the 3 days of testing under all experimental conditions, whereas the low IQ children in Groups I and II actually exhibit a decrement in performance from Day 2 to Day 3. Thus, it would appear that for low IQ children, consistency of problem type may be a necessary condition for learning set acquisition (as was the case in Group III), whereas the particular programming sequence may not influence the brighter children at all.

Experiment II

An analysis of variance conducted on the mean trials to criterion scores of the control and experimental groups on the first discrimination problem indicated no significant differences. The possibility exists, however, that retention effects might be observable in terms of a faster rate of positive

transfer. Consequently, the degree of improvement exhibited by each *S* from Problem 1 to Problem 3 was analyzed. These scores are presented in Table 2.

The analysis of variance conducted on these difference scores yielded a treatments effect significant at the .05 level ($F_{1,44} = 4.47$), indicating that the experimental group showed more improvement than the control *Ss* (an improvement of 44% compared with a decrement of 7%). Thus, *Ss* that had been previously exposed to discrimination training showed greater transfer. In addition, the Treatment $\times$ IQ interaction was significant ($F_{1,44} = 5.61, p < .05$), indicating a differential pattern of improvement for the high and low IQ *Ss* in the two conditions. For the experimental group, the high IQ *Ss* showed more improvement than the low IQ *Ss*, whereas the opposite was true in the control group.

TABLE 2
MEAN TRIALS TO CRITERION ON PROBLEMS
ADMINISTERED 6 MONTHS AFTER
ORIGINAL TESTING

Group	Problem		Difference
	1	3	
Experimental			
High IQ	10.18	5.36	4.82
Low IQ	7.92	4.75	3.17
Control			
High IQ	5.75	9.92	-4.17
Low IQ	10.92	8.00	2.92

DISCUSSION

The findings of the present investigation indicate that the type and sequence of discrimination problems employed may influence the rate of learning set acquisition. This effect is not a simple one, however, since the various groups differed only in the shape of the learning curves, and not in their performance on the criterion.

ion set of problems. The absence of a general treatment effect with regard to the criterion problems is somewhat puzzling. There are at least two possible explanations for this negative finding. The first is that the range of problem difficulty may not have been broad enough to accurately reveal transfer differences. In the present study discrimination difficulty was determined in advance on the basis of a theoretical model based upon the number of varying stimulus attributes. It is conceivable, however, that the children employed were not attending equally to all stimulus dimensions. It would seem advisable for future work in this area to assess difficulty level empirically in advance. The results of the present investigation suggest that discrimination difficulty level may be a function of intelligence as well as stimulus complexity. Another possible explanation for the lack of criterion differences is that the number of training trials was too large and thus may have masked any differences attributable to program sequence. It is possible that a smaller number of practice trials in each treatment group would reveal differences in programming efficiency. Future work is indicated to test this possibility. The major factor influencing performance in this study appeared to be amount of training received, since all three training groups were superior to a control group that did not receive prior training. Since Ss were drawn from a population generally regarded as deficient in learning experiences, it is interesting to note that the training used in the present experiment did indeed elicit learning sets.

There is some evidence that the particular programming of the problems may be a more significant variable in the performance of low than in normal IQ children. For a group with

normal intelligence, the particular problem sequence employed in this study did not seem to influence rate of acquisition although the sequence variable did affect the performance of low IQ Ss. For this latter group, consistency of problem type appeared to be the most efficacious presentation. The implication of this Treatment $\times$ Trials $\times$ IQ interaction for educational practice is that more careful attention may have to be given to the training experiences of children who are most deficient in learning skills.

The issue of how intelligence level is related to learning set performance is not entirely clear. Although a number of investigators have found that differences in rate of learning set acquisition are positively associated with intelligence level (e.g., Ellis et al., 1962; Jensen, 1963; Kaufman & Peterson, 1958; Koch & Meyer, 1959), other investigators have not obtained this relationship (e.g., Wischner & O'Donnell, 1962). The findings of the present investigation suggest that this relationship may occur only under certain kinds of experimental conditions. In earlier studies IQ differences in learning set acquisition have been obtained when the discrimination task employed has been relatively difficult or the success criterion very high (e.g., 90% correct in the Kaufman & Peterson, 1958, and Koch & Meyer, 1959, studies). Where the success criterion has been less stringent (e.g., two consecutive choices in the Wischner & O'Donnell, 1962, study, or three consecutive choices in the present study), differences between IQ groups have been less pronounced. Future research is indicated to clarify the relation of learning set acquisition to intelligence under a wider variety of experimental procedures.

The present study supports the ex-

pectation that when learning sets are acquired, they are retained by the organism over relatively long periods of time. This retention effect, once again, is not a simple one. The results of the second experiment indicated that a group previously exposed to training did not differ from a control group on the first problem presented after a 6-month interval. When improvement over a series of three problems was considered, however, the experimental group exhibited significantly more improvement. This suggests that the effect is not one of immediate memory, but rather one of enhanced speed of learning. Clearly, future investigations are needed which systematically vary the length of the retention interval and the type of original training in order to discover the most efficacious combinations.

REFERENCES

- BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychological Review*, 1959, **66**, 278-296.
- BRUNER, J. S. The course of cognitive growth. *American Psychologist*, 1964, **19**, 1-15.
- ELLIS, N. R., GIRARDEAU, F. L., & PRYER, M. W. Analysis of learning sets in normal and severely defective humans. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 860-865.
- HARLOW, H. The formation of learning sets. *Psychological Review*, 1949, **56**, 51-65.
- HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959.
- HARLOW, H. F., HARLOW, M. K., RUEPING, R. R., & MASON, W. A. Performance of infant rhesus monkeys on discrimination learning, delayed response and discrimination learning set. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 113-121.
- HAYES, K. J., THOMPSON, R., & HAYES, C. Discrimination learning sets in chimpanzees. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 99-104.
- JENSEN, A. Learning ability in retarded, average and gifted children. *Merrill-Palmer Quarterly*, 1963, **9**, 123-140.
- KAUFMAN, M. E., & PETERSON, W. M. Acquisition of a learning set by normal and mentally retarded children. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 619-621.
- KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology*, 1959, **58**, 56-60.
- KOCH, M. B., & MEYER, D. R. A relationship of mental age to learning set formation in the preschool child. *Journal of Comparative and Physiological Psychology*, 1959, **52**, 387-389.
- KORONAKOS, C., & ARNOLD, W. J. The formation of learning sets in rats. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 11-14.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin, 1956.
- MILES, R. C. Learning set formation in the squirrel monkey. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 356-357.
- MILES, R. C., & MEYER, D. R. Learning sets in marmosets. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 219-222.
- SHEPARD, W. O. Learning set in preschool children. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 15-17.
- WARREN, J. M., & BARON, A. The formation of learning sets by cats. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 227-231.
- WISCHNER, G. J., & O'DONNELL, P. Concurrent learning set formation in normal and retarded children. *Journal of Comparative and Physiological Psychology*, 1962, **56**, 735-739.

(Received July 4, 1966)

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 58

October 1967

Diary No.

Date

File No.

Bureau Ednl. Psy. Research

Number 5

CONTENTS

- Lecture versus Participation in the Health Training of Peace Corps Volunteers...* 259
JEAN S. KERRICK, VIRGINIA A. CLARK, AND DONALD T. RICE
- Learning from Prose Material: Length of Passage, Knowledge of Results, and Position of Questions...* 266
LAWRENCE T. FRASE
- Developmental Changes in the Recognition of Beginning Segments of English Words...* 273
ROBERT F. STANNERS AND DOLORES H. SOTO
- Pictorial and Verbal Factors in the Efficient Learning of Paired Associates...* 278
WILLIAM D. ROHWER, JR., STEVE LYNCH, JOEL R. LEVIN, AND NANCY SUZUKI
- Dogmatism and Examination Performance...* 285
B. JACK WHITE AND RICHARD D. ALTER
- A Twenty-College Study of Student $\times$ College Interaction Using TAPE (Transactional Analysis of Personality and Environment): Rationale, Reliability, and Validity...* 290
LAWRENCE A. PERVIN
- Differential Memory for Picture and Word Stimuli...* 303
JOSEPH R. JENKINS, DANIEL C. NEALE, AND STANLEY L. DENO
- Effects of Conceptual Similarity on Serial Learning and Retention by Retardates...* 308
ALFRED A. BAUMEISTER AND JUDITH GUFFIN
- Individual Differences in Memory Span with and without Activity Intervening between Presentation and Recall...* 311
ARTHUR E. WHIMBEY AND SANDRA L. LEIBLUM

LECTURE VERSUS PARTICIPATION IN THE HEALTH TRAINING OF PEACE CORPS VOLUNTEERS¹

JEAN S. KERRICK, VIRGINIA A. CLARK, AND DONALD T. RICE

School of Public Health, University of California, Los Angeles

A 2 × 2 factorial design with lecture versus participation as 1 variable and teaching team as the other was used to assess the health training of 272 Peace Corps volunteers. Before-after training tests were given for knowledge, attitudes, and beliefs about specific illnesses, and intention to follow recommended personal and community health behaviors. Results indicate that, with the specific task of health training, participation is not more effective than lecture, although teachers or teaching teams do differ in effectiveness. Further, if students do have a preference, it is for the lecture method.

When Peace Corps volunteers move from this country to overseas assignments, they move from an environment where the major diseases are those of old age to societies where the major diseases are still communicable diseases avoided only by behavior.

Health training, then, is only partially designed to impart knowledge. A major task is that of changing attitudes and intended health behavior. Volunteers must be persuaded to perform tasks which are annoying, repetitive, unpleasant, and sometimes embarrassing. For example, in some areas they must boil all water, even for bathing, they must not eat lettuce, ice cream, or other suspect foods, and they must often refuse to eat foods offered them by their hosts.

Two major lines of research lead to the present study. For more than 40 years, educators have been experimenting with lecture versus participation as classroom methods. For more than 20 years, psychologists and others have been concerned with

the two methods for short-term persuasion.

Studies as early as 1928 (Spence, 1928), and as late as 1954 (Ruja, 1954), indicate that, in the classroom, the lecture method is either superior to participation methods in imparting knowledge (as measured on multiple-choice information tests), or it is not significantly different from participation or discussion methods.

On the other hand, studies during the 1940s by Lewin (1947) and his associates clearly demonstrated the superiority of discussion or participation methods in changing reported behavior. While major variables in the early studies were confounded, Bennett (1955) clarified the issue. She concluded that the methods, *per se*, did not differ in effectiveness, but that the *decision* to act is the effective variable in increasing the probability of both reported and actual behavior.

The present study extends both lines of previous research. It extends the study of classroom learning by using a counterbalanced design to determine the effectiveness of the teacher under both methods, and to test the interaction between teacher and method. Second, it extends the time-span normally used in the studies of persuasive effect.

More important, however, the study focuses on a kind of behavior

¹This study was supported by Peace Corps Contract PC-(W)-381. Computing assistance was obtained from the Health Sciences Computing Facility, sponsored by National Institutes of Health Grant FR-3. The authors would like to thank Onelio Clark for programming assistance. The first author is now at Bolt Beranek and Newman Inc., Los Angeles.

unlike those possible in the usual experiment. In the earlier studies there were only minor unpleasant consequences of recommended behaviors. Even studies of risk-taking behavior (Bem, Wallach, & Kogan, 1965) use unpleasant potential consequences such as nausea and dizziness for only an hours' time. These studies have suggested that participation methods may be harmful when the consequences of behavior are unpleasant. But such minor unpleasant consequences can hardly be compared with potential effects of malaria, amebic dysentery, hookworm, or schistosomiasis.

The present study, then, examines lecture versus participation method in a classroom setting where the desired result is the extensive modification of behavior to reduce the risk of realistic illness with highly unpleasant consequences.

DESIGN AND PROCEDURE

Course Content

The first half of the 20-hour health-training course was focused on communicable disease theory; the classification, mode of spread of infectious diseases, immunity, and behavioral defenses against disease were studied.

Diseases prevalent in the host country, food- and water-borne diseases, respiratory, skin, zoonotic and insect-borne diseases, were described in relation to sanitation, environment, and recommended personal hygiene.

The second segment of the course covered problems of culture and health, community health education, school health and education, and community action related to health.

All instructors in the training course were given a standard outline of content and a list of teaching aids such as films, tapes, and readings. Each student was given a copy of the Peace Corps Health Manual, outlining the major illnesses and symptoms and treatments of prevalent diseases.

Experimental Design

The present study employed a 2×2 factorial design with lecture method and

participation method of teaching as one variable, and teaching team as the other.

Each teaching team was composed of a physician to teach personal health and a health educator to teach community health. Physicians and health educators were selected because they were experienced in both teaching methods, and were arbitrarily paired. Each team taught two sections, one by the lecture method, and the other by the participation method.

In lecture sections, instructors were asked to limit class questions and communication among students to 10 minutes. In participation sections, formal lecture was limited to 15 to 20 minutes per session, with 30 to 35 minutes of student-to-student interaction, emphasizing group discussion, buzz groups, role-playing situations, panels, and so forth.

Subjects (Ss) were 323 Peace Corps trainees destined for Ethiopia as teachers in secondary and higher elementary schools. For each of two daily periods available for health training, 50 Ss were randomly assigned to the participation section and the remainder were put into the lecture group. Of the original Ss, 293 completed training and 21 failed to complete the tests for the present study. The results are thus based on 272 trainees.

Measurements

Before and after health training, all trainees were given tests of factual knowledge, attitudes and beliefs about specific illnesses, and intention to follow recommended personal and community health behaviors.

The knowledge test consisted of 24 multiple-choice items and 10 matching questions. For the most part, information items had been used in earlier Peace Corps health training courses at the University of California, Los Angeles.

A previously constructed form of the semantic differential was used to assess attitudes toward illness (which we have called "severity"), and belief in the illness' avoidability or unavailability,² the two major factors isolated in a study of an illness-form of the differential. The illness test obtained responses to 25 illnesses likely to occur overseas.

²Our distinction between attitude and belief is the same that has been made by previous authors such as Krech and Crutchfield (1948), and Allport (1954), and operationally specified by Fishbein and Raven (1962).

A special form of the differential was constructed to determine expressed likelihood or intention to follow the major recommended health-related behaviors. Twenty-seven behaviors were included in the test. Expression of high "likelihood" of following behaviors appears directly comparable with measures of private commitment to behavior (Bennett, 1955).

A course-reaction questionnaire was given all four sections at the end of health training. The questionnaire, based on comments from previous volunteers, called for ratings of health training. Five of the eleven items were reversed, so that approximately half of the items were phrased as favorable, half unfavorable. Most items permitted "yes-no" answers ("Do you feel that you were given enough specific information?"), although two permitted three alternatives ("In general, how adequately did the lectures convey the information?" and "How helpful was the student participation in learning the health material?" "Very, somewhat, not very."). Five items related to material, six to teaching.

Personal health and community health segments of the course were rated separately on each item.

In summary, the following information was obtained before and after the health training course: knowledge of illness and disease, attitudes toward selected illnesses, beliefs in the avoidability or unavailability of those illnesses, and the extent to which the trainee saw himself as likely to carry out personal and community health-related behaviors. After training, course reaction was determined.

RESULTS

All data were run through standard screening programs giving posttest data plotted against pretest data, regression lines, and correlation coefficients. In most cases the randomization procedure adequately equated pretest measures, but the correlations between pre- and post-measures were high. Thus, the major analytic tool was a two-way (team versus method) nonorthogonal analysis of covariance, with pretest scores controlled. Table 1 gives the means for each group on all tests.

Knowledge Gain

The total group averaged 11.1 points higher on the posttest, Exam 2, than on the pretest, Exam 1, a significant overall difference (by *t*-test, $p < .05$). Thus, the total group did show significant information gain after the course. The covariance analysis indicated that teaching team had a significant effect ($p < .05$) while method did not. The groups taught by Team 2, on the average, learned more than those taught by Team 1. But neither the teaching method nor the interaction of Team $\times$ Method produced a significant difference. Team 2, then, was superior, regardless of method. The covariate (pretest score), as expected from the high correlation between the pre- and posttest, was significant, indicating that those who initially scored highest on the knowledge test tended to score highest on the posttest.

Attitude Change

Positive scores indicate illnesses which are "good" and "mild," and negative scores indicate illnesses which are "bad" and "severe."³ The group as a whole did not show significant attitude change, but the analysis of covariance did show a significant difference by teaching method ($p < .02$). Here, there was no difference between teaching teams and no Teams $\times$ Method interaction; but again, as expected, the covariate (initial attitudes toward illness) was significant. In general, illnesses tended to be viewed as bad and severe, but classes taught by the

³ The figures reported here are essentially mean summary scores over individuals and illnesses, using a procedure of weighting each scale response on the semantic differential according to its component loading on the attitudinal dimension, then weighting each illness according to its relative "goodness" or "badness."

TABLE 1
MEANS FOR EACH GROUP BEFORE AND AFTER TRAINING ON ALL TESTS

Item	Team 1			Team 2			Totals	
	Exam 1	Exam 2	r	Exam 1	Exam 2	r	Exam 1	Exam 2
Knowledge scores								
Participation	51.0	59.9	.80	51.1	62.3	.66	51.0	61.0
Lecture	52.3	62.2	.61	50.3	63.0	.69	51.1	62.7
Totals	51.8	61.3		50.5	62.8			
Attitude change (severity of illness)								
Participation	-65.5	-55.2	.67	-70.1	-56.1	.65	-67.6	-55.6
Lecture	-68.5	-75.1	.50	-66.2	-65.0	.57	-67.1	-69.0
Totals	-67.3	-67.3		-67.2	-62.8			
Belief change (avoidability of illness)								
Participation	-246.1	-252.6	.58	-225.8	-236.4	.80	-237.0	-245.4
Lecture	-237.0	-247.0	.67	-240.1	-251.5	.60	-238.9	-249.7
Totals	-240.6	-249.2		-236.5	-247.7			
Expressed likelihood of follow- ing recommended personal behaviors								
Participation	47.1	71.9	.45	47.9	66.7	.56	47.5	69.6
Lecture	47.5	73.5	.64	48.4	62.7	.52	48.0	66.9
Totals	47.3	72.9		48.3	63.7			
Likelihood of engaging in com- munity health behavior								
Participation	-22.4	-23.1	.68	-11.7	-12.8	.35	-17.6	-18.5
Lecture	-27.4	-32.1	.60	-21.4	-25.1	.71	-23.8	-27.8
Totals	-25.4	-28.6		-19.0	-22.0			

Note.—For Team 1, Knowledge Scores, Participation Group, $N = 46$, Lecture Group, $N = 74$; for all other items, Participation Group, $N = 47$, Lecture Group, $N = 73$. For Team 2, Knowledge Scores, Participation Group, $N = 38$, Lecture Group, $N = 114$; for all other items, Participation Group, $N = 38$, Lecture Group, $N = 113$.

participation method—either by Team 1 or by Team 2—believed illnesses to be less severe after training than before. In general, then, participation groups moved in the direction of more favorable attitudes toward illness, but it should be noted that a "favorable" attitude toward illness may not lead to illness-preventive behavior.

Belief Change

Table 1 presents changes in belief about the avoidability (positive scores) or unavailability (negative scores) of illnesses.⁴

⁴These scores were derived in the same manner as the attitude scores, using the component loadings on the avoidability

For the total group, illnesses are viewed as more avoidable after taking the course than before taking it. In the covariance analysis, however, only the covariate was significant. Thus, there was a significant correlation between pre- and posttest scores, but there was no difference in effecting belief-change by teams, methods, or by the interaction between them.

Change in Commitment

In general, trainees reported an increased likelihood of following be-

component of judgment toward illness, rather than the evaluative or attitudinal component, and weighting illnesses by their relative avoidability or unavailability.

haviors related to personal health, and a decreased likelihood of engaging in community health activity.

For both analyses, the pretest score made a significant difference. For personal health behaviors, however, there were no other significant differences. Thus, either team, by either method, was equally likely to raise the commitment to follow recommended personal health behaviors.

When community health was considered, however, method was a significant source of variation. Thus, those in the participation groups initially saw themselves as unlikely to engage in community health behaviors, and remained relatively unchanged by the training course. The lecture groups also initially expressed a lack of likelihood of engaging in community health behaviors, but saw these behaviors as even more unlikely after the course. It appears that the participation method made little change in the level of commitment to community action, while the lecture method actually lowered it, or conversely, increased a negative commitment.

Course Reaction

Guttman scales gave coefficients of reproducibility for material, presentation, and overall course reaction that were .93, .90, .87, indicating adequate reproducibility. Intercorrelations between Guttman scores for personal and community health segments indicate that students did make somewhat different judgments about the two segments, although the two were significantly correlated. Similarly, students distinguished between content and method. For personal health, content and presentation scores over individuals correlated .50, and for community health, .43.

Mean course reactions for the total group are approximately half

TABLE 2
MEAN COURSE REACTIONS BY TEACHING
TEAM AND METHOD: GUTTMAN SCALE
SCORES

	Personal health section		Community health section	
	Team 1	Team 2	Team 1	Team 2
Overall course reaction				
Participation	8.77	6.47	8.28	8.26
Lecture	6.85	4.88	8.15	7.93
Total	6.38		8.19	
Reaction to presentation				
Participation	6.32	4.26	5.68	5.76
Lecture	5.08	3.58	5.71	5.62
Total	4.61		5.74	
Reaction to material				
Participation	3.17	2.97	3.26	3.58
Lecture	2.75	2.34	3.15	3.18
Total	2.71		3.28	

Note.—High score indicates *unfavorable* response.

of the scale range (and medians also fall at about the midpoint of potential scores). Table 2 indicates that for personal health Team 2 was consistently superior to Team 1, according to students, and that lecture was consistently preferred to participation. While it is conceded that Guttman scores cannot be considered interval scales, some indication of the significance of these findings may be obtained from analysis of variance, keeping in mind that the statistical assumptions of normally distributed, continuous data are not met, although volunteers were randomly assigned to groups.

For personal health, by more than twice the *F*-ratio needed for the .05 level, both method ($F = 16.7$) and teaching team ($F = 24.7$) were significantly different, and there was no significant interaction between them.

It seems safe to say that both team and method produce different course reactions.

For the community health section, however, there were virtually no differences between either team ($F = .1$) or method ($F = .2$), and no F -ratio remotely approached significance. In short, neither the method nor the teaching team made a difference.

DISCUSSION

A two-way factorial design was used to test the effectiveness of lecture versus participation in health training of Peace Corps volunteers. Measures of knowledge gain, attitude change, belief change, and changes in commitment to action were tested by nonorthogonal analyses of covariance (controlling for pretraining scores) to determine the effect of teaching team, teaching method, and the interaction between them.

A review of the literature suggests participation is not superior in producing a knowledge gain. That hypothesis was supported by the present data. Lecture and participation groups gained the same amount of information, but one teaching team, using either method, was significantly better than the other. Thus, the teacher, not the method, appears to influence information gain.

Related to knowledge change, one aspect of belief about illness, its avoidability or unavailability, was affected by the training course. Volunteers viewed illness as more avoidable after training than before. This belief was not differentially affected by either teaching team or method, however.

A relatively recent review article (Stovall, 1958) predicts more attitude change by participation than by lecture. This, too, was supported by the present study. In general, the

participation groups saw illnesses as "better" or "milder" after the course than before it. With illnesses, however, it may well be that a favorable attitude toward illness is what one least wishes to encourage. One theoretical position (Rosenstock, Hochbaum, & Kegeles, 1960), suggests that health-related behavior will be less likely to occur when attitudes are "positive," and illnesses are seen as mild and less bad. In this instance, then, significant positive change may be interpreted as arguing against the use of the participation method where illnesses are being considered.

Given the specific sample, and the specific task of teaching knowledge about illness and behaviors which will prevent illness, it appears that the participation technique is not more effective in presenting knowledge (although the teacher or teaching team does differ) and it makes no difference in the one belief item measured.

Bennett (1955) observed that most of the differences in behavior change noted in lecture versus discussion experiments could be attributed to the individual's commitment to action, and that commitment was not directly related to teaching method. The present study controlled for commitment by asking all groups to rate their likelihood of following certain behaviors. Consistent with Bennett, the present study found that method did not differentially increase positive commitment toward either personal or community health behavior. The lecture method, however, significantly increased the *negative* commitment toward community health behaviors.

Perhaps the difference may be explained by Bennett's finding that discussion members tended to express slightly higher commitment, but no differences in actual behavior.

If increased knowledge, changing

attitudes, and commitment to recommended *personal behaviors* are our objectives, the lecture method is probably preferable. While some teachers are more effective in imparting knowledge, experienced teachers in the present study did equally well with either method in producing commitment to personal health behaviors.

It should be noted that in no case did a significant Teaching Team $\times$ Teaching Method interaction occur. This lack of interaction, however, may have resulted from the choice of instructors experienced in both methods. Under other conditions, significant interactions may be observed.

For course reaction, both team and method made a significant difference in the personal health segment. Team 2 was preferred to Team 1, and lecture was preferred to participation.

For the community health segments, neither team nor method made a significant difference.

In neither personal nor community health was there a significant interaction between team and method.

The present study suggests that where students are motivated and relatively interested, they prefer lecture to participation—perhaps because more material can be presented in a shorter amount of time. Spontaneous comments from students suggest that, in health courses,

they have too little information to utilize discussion, role-playing, or other participation methods effectively.

REFERENCES

- ALLPORT, G. W. *The nature of prejudice*. New York: Addison-Wesley, 1954.
- BEM, D. J., WALLACH, M. A., & KOGAN, N. Group decision making under risk of aversive consequences. *Journal of Personality and Social Psychology*, 1965, **1**, 453-460.
- BENNETT, E. B. Discussion, decision, commitment, and consensus in "group decision," *Human Relations*, 1955, **3**, 251-273.
- FISHEIN, M., & RAVEN, B. H. The AB scales: An operational definition of belief and attitude. *Human Relations*, 1962, **15**, 35-44.
- KRECH, D., & CRUTCHFIELD, R. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
- LEWIN, K. Group decision and social change. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 330-344.
- ROSENSTOCK, I. M., HOCHBAUM, G. M., & KEGELES, S. Determinants. Paper presented at the Golden Anniversary White House Conference on Children and Youth, Washington, March 1960.
- RUJA, H. Outcomes of lecture and discussion procedures in three college courses. *Journal of Experimental Education*, 1954, **22**, 385-394.
- SPENCE, R. B. Lecture and class discussion in teaching educational psychology. *Journal of Educational Psychology*, 1928, **19**, 452-462.
- STOVALL, T. F. Classroom methods. II: Lecture vs. discussion. *Phi Delta Kappan*, March, 1958, 255-258.

(Received July 8, 1966)

LEARNING FROM PROSE MATERIAL: LENGTH OF PASSAGE, KNOWLEDGE OF RESULTS, AND POSITION OF QUESTIONS

LAWRENCE T. FRASE
University of Massachusetts

72 college Ss read a 2000-word biographical prose passage. A $2 \times 3 \times 2$ factorial analysis assessed the effects of position of factual questions within the text, length of passage, and presence of knowledge of results. Posttest analysis focused upon (a) questions which had occurred during reading (retention questions), and (b) questions related to the section of the prose passage not tested by the retention questions (incidental questions). All 3 factors were significant for retention questions; only the position of questions was optimal for incidental questions. A moderate length passage was optimal for retention scores while scores on incidental questions tended to improve with longer passages. When questions occurred after passages both retention and incidental scores were high. Elicitation of "mathemagenic" behavior provided the most concise interpretation of the data.

The aim of the present study was to determine how factual questions might be used to improve retention of prose materials. Such questions, placed within a prose passage, might have specific effects, improving retention only on sections of the prose directly relevant to the question. The same questions might also have a more general effect, influencing the retention of prose information which is incidental to the questions asked.

Rothkopf (1965, 1966) has investigated issues raised by the general problem posed above. He found that questions can have both a specific and general facilitative effect upon retention of prose material. He also found that a simple instruction to read carefully facilitated retention in general. Rothkopf maintains that questions elicit a general class of responses, called "mathemagenic behaviors," which aid in the acquisition and retention of information. The "cybernetic" approach complements the view held by Rothkopf. Smith and Smith (1966) imply that questions may provide an internalized criterion for attending to the content of a

prose passage. According to this view, question-relevant prose content would elicit appropriate mathemagenic behaviors. Placing the question in front of the prose passage would be a necessary condition for this effect to occur. Rothkopf (1966) concludes from his data, however, that specific effects of question are obtained regardless of the location of questions. In the present study this problem was explored by placing factual questions either before or after passages to which they relate.

Another question explored by Rothkopf was the use of knowledge of results (KR). When questions come before related prose passages and knowledge of results is given, the prose material may interfere with the retention of the questions. In short, the reading of prose material, in this situation, may be viewed as a form of interpolated learning between questions. Berlyne (1966) has suggested the need for research in this area. Intuitively it seems that if a question occurs without KR, reading the relevant prose passage after the question should provide KR and hence reten-

tion test scores should be improved. Rothkopf's (1966) data indicated that this may not be the case. His results suggest the possibility of an interaction between KR and position of questions, but the nature of this interaction was not clear from his data. Rothkopf's major emphasis was on the comparison of his individual groups with each other and with a control group. The present study, rather than focusing upon the control condition, provided information concerning possible interaction of the variables studied.

A final variable of potential relevance is the amount of prose material interpolated between questions. Is there an optimal rate of questioning and does this rate differentially affect the retention of information which is relevant or incidental to the questions asked? Ausubel's (1962) view of meaningful verbal learning, contrary to current "active response" requirements of programmed instruction, suggests that retention may be improved if subjects (Ss) are allowed to read passages as a whole. Instead of spacing prose reading by interpolating questions between passages, Ss reading prose under whole conditions would be more likely to relate the facts contained in the passage meaningfully. An alternative view is that if questions occur in the prose material they should occur in close proximity to the relevant content. Programming, which uses test-like events, necessitates breaking the material into small chunks to facilitate associations between questions and content. In the present study, size of passage between questions was varied. The two types of posttest questions mentioned above (relevant and incidental) pertain to prose content in which relevant test items occur or do not occur. Perhaps the pro-

gramming approach and Ausubel's view are both correct, in which case the effect of length of material would differ for retention and incidental questions.

METHOD

Subjects and Experimental Design

The experiment was presented as a laboratory assignment to 79 students of introductory educational psychology.

A $2 \times 3 \times 2$ factorial design was used, the three factors being: (a) questions before or after the reading passages, (b) length of passage (10, 20, or 40 lines long), and (c) KR present or absent following the questions.

Stimulus Material

A continuous prose passage of biographical material on William James was selected from *Psychology: The Science of Mental Life*, by Miller (1962).¹ The passage was divided into 20 paragraphs of 10 lines each. The material was reproduced on mimeograph with each paragraph a separate sheet of 4×11 -inch paper. Two multiple-choice questions (5 alternatives) which required the recall of specific factual information (such as the number of children in the James family, or a course of study undertaken by William James at a particular time) were constructed for each paragraph. These two questions did not overlap in terms of their content. The 20 questions relating to the second part of the paragraphs were used as *retention questions* on the posttest. They occurred during the learning session (either before or after passages) and also on the posttest. The other set of 20 questions, relating to the first half of the paragraphs, was used to test for *incidental learning*. Pretesting indicated that the two sets of questions were at the same chance level of difficulty.

Procedure

The experiment was given on 2 consecutive days. The Ss were randomly assigned to one of the 12 groups for 1 of the 2 days. This meant that $\frac{1}{2}$ of the Ss under each

¹Permission for the experimental use of these copyrighted materials was kindly granted by the publishers, Harper & Row, Inc., 49 East 33rd Street, New York, N. Y. 10016.

TABLE 1

COMPARISON OF PERCENTAGES CORRECT
WITH DATA OF ROTHKOPF (1966)

	LBA	B	BA	A	AA	Control
Rothkopf						
Retention	78	65	78	63	82	29
Incidental	36	30	35	43	40	33
Fraser						
Retention	85	61	87	79	91	68
Incidental	52	54	53	73	70	61

Note.—LBA = all questions and answers were given, then *Ss* read passages (in the present study the 40-line length most nearly approaches this condition); B-questions occurred before each paragraph; BA = questions and answers given before each paragraph; A = questions given after each paragraph; AA = questions and answers given after each paragraph; Control = read prose passages. For the Rothkopf groups, $N = 20$; for Fraser, $N = 18$ for all groups except LBA ($N = 6$) and Control ($N = 7$).

condition were run on the first day and $\frac{1}{2}$ on the second day. Two dependent measures were obtained immediately after the reading task; (a) the number of correct responses to the 20 questions which had occurred in the learning task (retention questions), and (b) the number of correct responses to the 20 questions covering material not questioned in the reading task (incidental questions). The retention and incidental test scores were analyzed separately. They might have been included as a dimension in a repeated measures design, but the resulting confounding of some of the conditions with KR makes interpretation ambiguous.

A control group of seven *Ss* randomly assigned from the same class was run along with the other *Ss* to indicate a base-line level of responding by simply reading through the prose material without the test questions.

The reading material took the form of a stack of unbound pages. On one sheet of paper *S* found a paragraph of prose material, on another sheet a question over that material, and on another sheet KR was given by repeating the number and content of the correct alternative along with the stem of the question. The sequence of paragraph-question-KR that a particular *S* saw depended upon the experimental group to

which he was assigned. As an example, one *S* might read the following six pages: (a) Paragraph 1, (b) Question 1, (c) Answer 1, (d) Paragraph 2, (e) Question 2, (f) Answer 2. Another *S* might see the same material in a different sequence: (a) Question 1, (b) Answer 1, (c) Paragraph 1, etc., or, if the size of the material were varied; (a) Paragraph 1, (b) Paragraph 2, (c) Question 1, (d) Answer 1, (e) Question 2, (f) Answer 2. If KR was given, the answer always followed the question to which it related.

The *Ss* were directed to seats on one side of tables (in a laboratory classroom) upon which the reading materials had previously been placed. They were cautioned not to look through the materials until told to do so. After *Ss* had been seated, the experimenter read standard instructions stating that this was an experiment to see how much people can learn from reading material. The *Ss* were told to read each page of their material (turning each page face down after they read it) and not to turn back once they had read a page. They were told that when they encountered a question in the text they were to try and get the answer before going on to the next page. They were told that a test would be found at the end of the reading material, and that they were to complete the test immediately after the reading task.

The *Ss* were also informed that some of the materials were shorter than others and hence some *Ss* might finish sooner than others. The procedure was again explained. The experimenter insured that the instructions were understood before proceeding with the experiment.

RESULTS

Comparison of Two Studies

Table 1 presents data from groups in the present study which are comparable to groups used in Rothkopf's study.

Rothkopf used a 5,000 word passage from Rachael Carson's *The Sea Around Us*, and fill-in questions rather than multiple-choice items. Hence, the two studies differ in terms of size and content of material, and response mode. In spite of these differences Table 1 reveals that the relative standing of the groups was quite

TABLE 2
SUMMARY OF MAIN EFFECTS

	Retention questions			Incidental questions		
	<i>M</i>	<i>F</i>	<i>df</i>	<i>M</i>	<i>F</i>	<i>df</i>
Question position		16.10	1/60**		22.00	1/60**
Before	15.14			10.67		
After	17.00			14.28		
Length of passage		3.94	2/60*			
10	15.96			11.33		
20	16.92			12.96		
40	15.33			13.13		
KR		52.20	1/60**			
Present	17.78			12.28		
Absent	14.36			12.66		

* $p < .02$.

** $p < .001$.

similar across the two studies. The rank-order correlation of the means of the two studies is significant at the .01 level, attesting to the generality of Rothkopf's findings. The difference in length of material and response mode accounts for the higher average scores in the present study.

The reader is referred to the note under Table 1 for an explanation of the group labels. It can be seen that in both studies the group which received questions (without KR) before prose passages (Group B) performed somewhat below the Control Group mean on incidental questions. The difference between the Control and B Group means was not statistically significant, but the consistency across studies suggests that there may actually be a depressing effect of factual prequestioning upon incidental learning which the test items were not sensitive enough to detect.

Retention Questions

There was a significant main effect of all three factors upon retention questions (Table 2). Presenting questions before paragraphs had the least facilitating effect upon re-

tention. There was a significant interaction between the position of questions and whether KR was available or not ($F = 4.42$, $df = 1/60$, $p < .05$). When KR was available, the means for the before-after position of questions were 17.33 and 18.22; when KR was not available, the means were 12.94 and 15.78, respectively. In short, the position of the retention questions did not make much difference if KR was provided. If KR was not provided, then questions were more effective following the passages. This result was somewhat surprising. It was anticipated that Ss would obtain KR (when it was not given with questions) providing that the questions occurred before the prose paragraphs. Contrary to the cybernetic hypothesis, Ss did not seem to internalize questions and focus their reading behaviors on question-relevant content.

Analysis of the length of passages reveals that the 20-line passage was optimal for facilitating retention. Presumably, if passages are too long, associations between questions and content are difficult to establish. If passages are too short, continuity

among prose content is broken. This result is especially interesting because the effect of length upon incidental questions was quite different (see below).

The result of including KR upon retention items confirms the expectation that if students are given the answers to questions which will occur on an immediate posttest, their performance will be relatively high on that test.

Incidental Questions

Table 2 reveals one significant effect for incidental questions. Placing retention questions after prose passages tended to increase the amount of information acquired for other portions of the material. A similar effect was obtained with the retention questions (contrary to Rothkopf's nonsignificant findings in this regard), but the difference between the before-after position was even more pronounced for incidental questions. The results clearly indicate that, if questions are to have a general facilitative effect upon retention of prose material, then the questions should occur after passages.

The differences between levels of passage length were not significant for incidental questions. In contrast to the significant curvilinear relationship for retention questions across length of passage, there was a gradual improvement in scores on incidental questions with the larger passages. This trend, in accordance with Ausubel's (1962) view, implies that reading larger passages may be an optimal procedure when relevant guidance, in the form of short tests, does not occur in the prose material. Hence, support is given to both the small step and whole reading approaches depending upon whether one emphasizes re-

tention of selected facts or retention of the total prose passage.

DISCUSSION

The results of the present study support the data of Rothkopf (1965, 1966) which indicate that short tests included with prose passages facilitate retention of specific and incidental information. Placing test questions after the prose passages was the optimal procedure for both specific and general retention. There are two relevant hypotheses which relate to this result. The first hypothesis states that placing test questions after passages requires Ss to review implicitly content which has just been read; therefore retention of the preceding material is facilitated. This hypothesis asserts that questions work in a *backward* manner, organizing and repeating previous prose content. Ausubel, Schpoont, and Cukier (1957) found that asking Ss to remember course material after they had read the material did not facilitate retention. It is difficult to see how Ss could make use of questions in a review capacity unless they were able to reproduce the preceding passage.

The alternative hypothesis asserts that questions act in a *forward* manner optimizing "mathemagenic" behaviors on passages following the questions. Questions which occur before prose passages (and are relevant to those passages) tend to limit the general facilitating effects of test-like events since they relate to specific content within the passages. In effect, placing questions after passages (as review) alters their function. The questions now provide a general test-taking orientation applicable to the paragraphs which follow. The forward hypothesis accounts for several

facts. First, Rothkopf (1966) obtained relatively high performance on incidental questions from a group which received only a general instruction to read carefully. Such an instruction, although not frequent enough to maintain attentive responses, should provide some facilitation. Second, when questions were placed before paragraphs in the present study, retention of incidental information was relatively low. Factual prequestioning evidently limits the range of attentive behaviors. Finally, recall of retention questions (when they occurred before passages) was high relative to incidental questions—a repetition effect—but retention was even higher when the questions followed paragraphs (a repetition and general mathemagenic effect).

A critical issue concerning the forward and backward hypotheses is the degree to which retention of the information in *one* paragraph is facilitated by placing a relevant question before it as opposed to an irrelevant question. In the present study and in Rothkopf's (1966), questions which occurred after a relevant paragraph also acted as irrelevant questions preceding the next paragraph. Further experimentation, in which a paragraph is either preceded or followed by a question (but not both), might contrast the relative strength of the repetition-review and attentive effects of questions.

According to the cybernetic hypothesis, questions must occur before prose passages to focus effectively reading behaviors. If prequestioning were a potent variable in terms of controlling inspection behaviors, then lack of KR (with prequestioning) should have resulted in poorer retention of incidental prose material,

which it did not. If focusing did occur, then it was not the result of epistemic behaviors (Berlyne, 1966), that is, a search for KR, but the result of a more general orientation induced by the factual questions. Experiments should be conducted to determine the range of questions which produce general or specific facilitating effects. In terms of the previous discussion, an alternative means of inducing general facilitative effects would be to manipulate the function of questions by changing their position.

To summarize, the present study extended the findings of Rothkopf (1966). Data indicate that test questions induce appropriate attentive responses when placed after the prose passages. Optimal spacing of test items in the prose material may differ for retention of test-relevant as opposed to incidental prose content. An important unresolved issue concerns the relative strength of forward and backward effects of questions. Backward effects relate to review and repetitive effects of relevant questions which follow paragraphs. Forward effects, apparently quite potent in the present study, refer to general orientation or "mathemagenic" responses which occur when unrelated questions precede passages. Another forward effect, suggested by the cybernetic hypothesis, concerns specific orientations when related questions precede passages. The latter did not appear as a strong factor in this study.

REFERENCES

- AUSUBEL, D. P. A subsumption theory of meaningful verbal learning and retention. *Journal of General Psychology*, 1962, 66, 213-214.
- AUSUBEL, D. P., SCHPOONT, S., & CUKIER, L. The influence of intention on the reten-

- tion of school materials. *Journal of Educational Psychology*, 1957, **48**, 87-92.
- BERLYNE, D. E. Conditions of prequestioning and retention of meaningful material. *Journal of Educational Psychology*, 1966, **57**, 128-132.
- MILLER, G. A. *Psychology: The science of mental life*. New York: Harper & Row, 1962.
- ROTHKOFF, E. Z. Some theoretical and experimental approaches to problems in written instruction. In J. Krumboltz (Ed.), *Learning and the educational process*. Chicago: Rand McNally, 1965. Pp. 193-221.
- ROTHKOFF, E. Z. Learning from written instructive material: An exploration of the control of inspection behavior by testlike events. *American Educational Research Journal*, 1966, **3**, 241-249.
- SMITH, K. U., & SMITH, M. F. *Cybernetic principles of learning and educational design*. New York: Holt, Rinehart & Winston, 1966.

(Received September 8, 1966)

DEVELOPMENTAL CHANGES IN THE RECOGNITION OF BEGINNING SEGMENTS OF ENGLISH WORDS

ROBERT F. STANNERS¹ AND DOLORES H. SOTO²

Washburn University

This experiment was designed to investigate the ability of English-speaking children to distinguish, by means of a visual-reading method, among highly frequent initial 3-letter word beginnings (consonant-consonant-vowel), infrequent word beginnings, and CCVs which never occur in English. A modified method of paired-comparisons was used to present the CCVs to 68 3rd graders, 70 6th graders, and 71 9th graders. The discriminatory ability for all grades was significantly above chance, $p < .0001$. The 6th and 9th graders were not significantly different from each other, but both were better than the 3rd graders, $p < .001$. It appeared that discriminatory ability for the materials was already well begun by the end of the 3rd grade, virtually maximal by the end of the 6th grade, and did not differ with respect to sex.

Linguists Chomsky and Halle (1965) have suggested that from the speech around him a child learns a grammar which allows him to form the admissible phonological combinations of certain letters but not others. In English, for example, words do not begin with *sb*, *ny*, *dz*, or many other combinations of letters which may occur in other languages. Church (1961) reports that even by age 2 or 2½, children have learned the rules for combining consonants and vowels well enough to be able to detect alien combinations. Carroll (1960) is of the opinion that the overall frequency with which language items occur in the speech heard by the child has a positive relation to the developmental sequence in which the items are learned. As Chomsky and Halle (1965) put it, the child hears *brik* (brick) but not *blik* or *bnik*. Since he hears words like *black* and *blanket*, he builds a grammar that contains the rule that *blik* is admissible, even though it is not meaningful; but, since he hears

no words beginning with *bn*, he will reject *bnik* as a possible English word.

Two studies (Brown & Hildum, 1956; Messer³) have shown that correct aural identification of verbal units is a function of their probability of occurrence in English. Brown and Hildum (1956) compared the ability of English speakers to identify nonsense words having unlawful (that is, not following the phonological rules of English) initial clusters with their ability to identify lawful nonsense words and low-frequency English words. The three types of words were presented aurally and adult subjects (Ss) were asked to spell the utterance. It was found that more low-frequency English words were correctly identified than lawful nonsense words, and there were more correct identifications of the lawful nonsense words than of the items with unlawful initial clusters. Messer performed his study with 4-year-old children to determine whether they could distinguish phonologically lawful non-

¹ Now at Oklahoma State University.

² Undergraduate research participant under the National Science Foundation program.

³ A brief report of this study appears on page 61 of the 1964-1965 Fifth Annual Report from the Center for Cognitive Studies, William James Center for the Behavioral Sciences, Harvard University, Cambridge, Massachusetts.

TABLE 1
CCVs WITH IDENTIFICATION NUMBERS, FRE-
QUENCIES OF OCCURRENCE, AND SCALE
SCORES FOR THE VARIOUS GRADES

Identifica- tion number	CCV	Fre- quency of occurrence	Scale score		
			Third	Sixth	Ninth
1	WHI	125	2.349	3.863	3.196
2	STA	91	2.536	3.511	3.406
3	THI	234	2.545	3.549	3.266
4	CHA	62	2.215	3.647	3.315
5	PLA	70	2.570	4.028	3.553
6	PRO	53	2.137	4.037	3.763
7	THO	52	2.182	4.013	3.393
8	DWE	1	1.225	1.528	1.727
9	BLE	1	1.890	2.893	2.758
10	SCU	1	1.866	2.374	2.495
11	SNU	1	1.797	2.047	2.344
12	SWO	1	1.685	2.422	2.597
13	WRA	1	2.041	2.791	2.596
14	WRE	1	2.099	2.819	2.756
15	SBA	0	1.260	1.586	1.933
16	FWO	0	1.042	1.352	1.500
17	MBO	0	0.862	1.483	1.475
18	RWA	0	0.847	1.393	1.158
19	NWE	0	0.916	1.262	1.226
20	FPA	0	0.986	1.392	1.089
21	DZO	0	1.000	1.000	1.000

sense words from unlawful nonsense words. The material was presented aurally and the results showed that about 70% of the time the children were able to distinguish the lawful nonsense words from the unlawful ones. Gibson, Osser, and Pick (1963) presented pronounceable and unpronounceable trigrams tachistoscopically to first- and third-grade children and found that units which fit the rules of English, in terms of pronounceability, were recognized more accurately than those which did not. In a previous study (Gibson, Pick, Osser, & Hammond, 1962, p. 555) it was hypothesized that "reading consists of decoding graphic material to the phonemic patterns of spoken language which have already been mastered when reading is begun" and that the observation of grapheme-phoneme correspondences leads to skilled recognition.

The present study was concerned with the ability of English-speaking children (7½-15½ years) to distinguish highly frequent initial 3-

letter word beginnings from those of low frequency, and from those which never occur in English. Samples of third-, sixth-, and ninth-grade children were used to determine if knowledge of English phonological rules, as indicated by a visual-reading method, is a developmental sequence beyond the age of 7½ or 8 years, and whether there are sex differences in the level of this knowledge.

METHOD

Subjects

The Ss were 209 public school children from two elementary schools and one junior high school in Shawnee County, Kansas. Approximately one-half of the third- and sixth-grade samples came from the same area as did the ninth-grade sample. The other one-half of the third- and sixth-grade samples came from another part of the county. As nearly as could be determined, all members of the total sample were of the same general middle-class background. All Ss spoke English as a native language and all were Caucasian except for two Ss who were Negro. Specifically, there were 68 third-grade children (34 boys ranging in age from 89-123 months, and 34 girls, 88-112 months of age); 70 sixth-grade children (41 boys, 137-162 months of age, and 29 girls, 137-160 months of age); and 71 ninth-grade children (29 boys, 173-189 months of age, and 42 girls, 172-185 months of age).

Materials

Three lists, each containing seven consonant-consonant-vowel (CCV) combinations, were constructed. One list was composed of the seven most frequently occurring initial (i.e., occurring in positions 1-3 of a word) CCVs obtained from the tables published by Mayzner, Tresselt, and Wolin (1965). Another list of initial CCVs which occurred only once in the sample was obtained from the same source. The third list of CCVs which could not occur under the phonological rules of English was obtained from Whorf's formula (Brown, 1965, p. 268). Table 1 presents the 21 CCVs and their frequencies of occurrence in a collection of 20,000 English words, 4-7 letters in length. Frequencies were calculated by summing over word lengths 4-7 for letter positions 1-3.

Procedure

The arrangement of the CCVs for presentation was accomplished by arbitrarily assigning a number between 1 and 7 as an identification number to each of the highly frequent (HF) CCVs; the numbers 8-14 to the infrequent (LF) CCVs; and the numbers 15-21 to the nonoccurring (N-O) CCVs. CCV 1 was paired with CCVs 8-12 and 15-19; CCV 2 with CCVs 13, 14, 8-10, and 20, 21, 15-17; CCV 3 with CCVs 11-14 and 8, and 18-21 and 15. This system was continued until each of the highly frequent CCVs had been paired 10 times with CCVs from the other two groups. The same system was used to pair the LF CCVs with the N-O CCVs. The system resulted in a total of 105 pairings, which was exactly one-half the number which would have been required if the usual method of paired-comparisons had been used. The CCV pairs were arranged in random order with the restriction that each reappearing CCV was as far as possible from its last appearance in the list. Each CCV was presented 10 times, and each time it was paired with a different CCV; however, there were no within-group pairings. Right-left placement of "correct" selections (i.e., the CCVs most frequent in English) within the entire list of 105 pairings was controlled with one-half of the correct choices occurring on the left, and the other one-half on the right. Sequential patterning of correct right-left responses was arranged randomly. The CCV pairs were presented to each *S* in a small booklet of 105 pages with one pairing on each page. All booklets were identical except that one-half of them were arranged in reverse order to control for practice and fatigue.

The *Ss* were run in groups in their respective classrooms during May, 1966. The session was untimed, with the administrator reading aloud the instructions and offering four visual examples of paired-comparisons which were similar to those in the materials. The *Ss* were asked to write their sex and birth date on the back of the booklet, and to underline the letter combination on each page which they had seen or heard most often as the beginning of an English word. All questions by *Ss* were answered before and during the administration.

RESULTS AND DISCUSSION

The first phase of the analysis was to test the mean proportion of judgments correct, that is, in conformity with the objective frequency count,

against a chance hypothesis, a proportion of .500. For the third grade the mean was .795, $t(67) = 16.52$, $p < .0001$; for the sixth grade the mean was .908, $t(69) = 44.27$, $p < .0001$; for the ninth grade the mean was .923, $t(70) = 58.12$, $p < .0001$.

Paired-comparison scores, presented in Table 1, were calculated for each of the CCVs by the incomplete matrix method (Torgerson, 1958, Ch. 9). A separate analysis was done for each sample of *Ss* from the three grades. A value of 1.00 was used as an arbitrary origin point for the scale. The means of each set of seven scores in a given frequency category are plotted by grades in Figure 1. A two-way analysis of variance was performed on these data, and all effects were found to be statistically highly significant. For grades, $F(2/36) = 129.42$, $p < .001$; for frequency, $F(2/18) = 110.20$, $p < .001$; and for the interaction of Grades $\times$ Frequency, $F(4/36) = 19.02$, $p < .001$. Figure 1 indicates that the significant interaction term may be attributed largely to the difference in the slopes of the trend lines of the sixth and ninth grades on the one hand, and the third grade on the other. The steeper slopes of the trend lines for the sixth and ninth grades, as compared to the third, indicate a higher level of discrimination among the CCVs in the judgment of familiarity. Because of the significant interaction effect and the relative closeness in the means of the sixth and ninth grades, as compared to the third, two-group comparisons were made among the means of the grades by Scheffé's test (Edwards, 1962). The results of the tests are presented in Table 2. The difference between the means of the sixth and the ninth grades in the HF category is a result of the sixth graders showing a slightly higher level of differentiation among the CCVs *within*

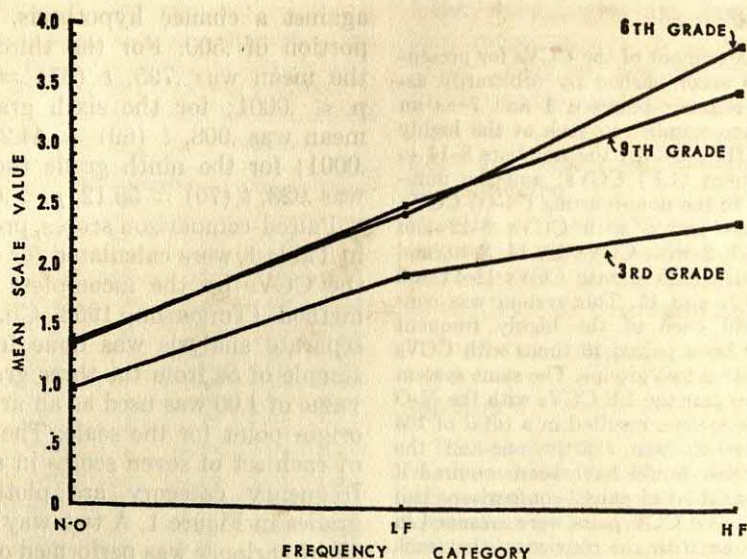


Fig. 1. Average familiarity scale score for the three grades as a function of CCV frequency.

the HF category than did the ninth graders. The sixth grade's proportion scores for the various pairings of the HF CCVs with CCVs from the other two categories were somewhat more variable than were those of the ninth grade. The greater variability indicated that the sixth graders perceived slightly larger differences in familiarity among the HF CCVs and is reflected in the small increase in the scale scores. All other significant differences were between the third

and sixth grade and the third and ninth grade. The tests of the mean proportions, the analysis of variance, and the multiple comparisons tests indicate that a fairly high degree of discriminatory capacity is learned by the end of the third grade and that the level of learning is virtually asymptotic by the end of the sixth grade.

The next analysis was concerned with differences in the mean proportion correct among the three grades by sex. A two-way analysis of variance for unequal cell frequencies (Winer, 1962, Ch. 5) was performed on the proportion data. For grades, $F(2/203) = 51.14$, $p < .001$; for sex, $F(1/203) = 0.00$; and for the interaction of Grades $\times$ Sex, $F(2/203) = 0.00$. As would be inferred from the value of the latter two terms, the differences between the means for boys and girls over the three grades were extremely small. For girls and boys, respectively, in the third grade the means were .793 and .798; in the sixth grade, .912 and .906; and

TABLE 2

VALUES OF F ASSOCIATED WITH MEAN COMPARISONS OF SCALE SCORES USING SHEFFE'S TEST

Comparison	Frequency Level		
	N-O	LF	HF
3rd vs. 6th	15.13	42.27	237.04
3rd vs. 9th	14.11	50.54	125.45
6th vs. 9th	0.02	0.37	17.60

Note.— $F = 13.86$ is needed for significance at the .01 level; $F = 25.94$ at the .001 level. N-O nonoccurring; LF = low frequency; HF = high frequency.

in the ninth grade, .923 and .923. The overall values for girls and boys were .876 and .876. The two-group differences among the means for the three grades were tested by Scheffé's test (Edwards, 1962). For the third grade versus the sixth, $F(2/135) = 44.02$, $p < .001$; for the third grade versus the ninth, $F(2/136) = 56.897$, $p < .001$; and for the sixth grade versus the ninth, $F(2/138) = .793$. Virtually all of the large main effect for grades which was found in the analysis of variance may be attributed to the difference between the sixth and ninth grades on the one hand, and the third grade on the other. This finding may be adduced in support of the conclusion that the development of discriminatory capacity for initial CCVs is largely complete by the end of the sixth grade. The generality of the conclusion is limited, of course, by the sample of materials used in the present study. The nearly identical performance for boys and girls is somewhat surprising in view of the common finding that girls are superior on verbal tasks (Anastasi, 1958, Ch. 14). Additionally, boys have been found to have a greater incidence of reading difficulties in grade school (Bentzen, 1963).

Since the familiarity scale scores for the CCVs could be useful in other research, the stability of the scores was checked by calculating the correlations among the three sets of scores from the three independent samples. For the third grade and sixth grade, $r = .94$; for the third and ninth, $r = .97$; and for the sixth

and ninth, $r = .97$. The rank-ordering of the CCVs, even within a frequency category, was very similar across the three grades.

REFERENCES

- ANASTASI, A. *Differential psychology*. New York: Macmillan, 1958.
- BENTZEN, F. Sex ratios in learning and behavior disorders. *American Journal of Orthopsychiatry*, 1963, **33**, 92-98.
- BROWN, R. *Social psychology*. New York: The Free Press, 1965.
- BROWN, R., & HILDUM, D. C. Expectancy and the perception of syllables. *Language*, 1956, **32**, 411-419.
- CARROLL, J. B. Language acquisition, bilingualism, and language change. In C. W. Harris (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1960. Pp. 744-752.
- CHOMSKY, N., & HALLE, M. Some controversial questions in phonological theory. *Journal of Linguistics*, 1965, **1**, 97-138.
- CHURCH, J. *Language and the discovery of reality*. New York: Random House, 1961.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1962.
- GIBSON, E. J., OSSER, H., & PICK, A. D. A study of the development of grapheme-phoneme correspondences. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 142-146.
- GIBSON, E. J., PICK, A., OSSER, H., & HAMMOND, M. The role of grapheme-phoneme correspondence in the perception of words. *American Journal of Psychology*, 1962, **75**, 554-570.
- MAYZNER, M. S., TRESSELT, M. E., & WOLIN, B. R. Tables of trigram frequency counts for various word length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1**, 13-24.
- TORGESON, W. S. *Theory and method in scaling*. New York: Wiley, 1958.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received September 12, 1966)

PICTORIAL AND VERBAL FACTORS IN THE EFFICIENT LEARNING OF PAIRED ASSOCIATES¹

WILLIAM D. ROHWER, JR., STEVE LYNCH, JOEL R. LEVIN, AND
NANCY SUZUKI

University of California, Berkeley

An experiment was designed to assess 2 methods of facilitating the acquisition of paired associates. Each of 96 3rd- and 96 6th-grade children was asked to learn a list of 24 pairs by a study-test method. The pairs were either pictures of objects or the printed names of those objects. As each pair was presented, the experimenter uttered 1 of 4 kinds of verbalization: (a) the names of the objects; (b) the names of the objects connected by a conjunction; (c) the names of the objects connected by a preposition; and, (d) the names of the objects connected by a verb. For both older and younger Ss, pictorial materials produced more efficient learning than printed materials and verb connectives uniformly facilitated acquisition. The relationship between connective form class and amount learned was such that verbs produced more learning than prepositions and prepositions produced more than conjunctions. The form of this relationship varied with materials and with grades.

Surely, one of the important tasks of pedagogy is to create conditions that produce efficient learning. Two possible ways of accomplishing this task, at least for the kinds of learning that conform to the paired-associate paradigm, are close at hand. The first derives from evidence reported in a series of studies of the effects of sentence contexts on the learning of noun pairs. The learning materials used have consisted of small objects (Jensen & Rohwer, 1963a), pictures of objects (Jensen & Rohwer, 1963b, 1965) and printed names of objects (Rohwer, in press; Rohwer & Lynch, 1966, in press; Rohwer, Shuell, & Levin, in press). The typical design of these studies contrasted the amount learned when the noun

pairs or object names were presented in the context of a sentence uttered either by the experimenter (*E*) or by the subject (*S*) or both, with the amount learned when no sentence context was provided. Uniformly, the sentence conditions produced more efficient learning by mentally retarded adults (Jensen & Rohwer, 1963a, 1963b), second-, fourth- and sixth-grade school children (Jensen & Rohwer, 1965), and fifth- and sixth-grade children (Rohwer & Lynch, 1966). Using a more finely-graded series of conditions, Rohwer (1966) found that the presentation of noun pairs in the context of an English string in which the nouns were connected by a verb or a preposition (e.g., The running COW chases the bouncing BALL; The running COW behind the bouncing BALL) produced more rapid learning than strings in which the pairs were connected by a conjunction (The running COW and the bouncing BALL). The reported as-

¹This work was supported, in part, by a contract with the United States Office of Education (OE6-10-273) through Project Literacy. The report was prepared at the Institute of Human Learning, which is supported by grants from the National Science Foundation and the National Institutes of Health.

sociation between gradations in learning efficiency and the form class of context connectives is in need of replication, but the general result is reliable: Sentence contexts facilitate the learning of noun pairs.

A second method that appears to have potential for promoting learning efficiency is that of presenting materials pictorially rather than in print. In connection with the question whether or not foreign-language words are learned more easily when associated with their environmental referents or when associated with native-language word equivalents, Wimer and Lambert (1959) found that nonsense syllables were learned faster as responses to objects than as responses to the printed names of the objects. Since frequently it is difficult to use many kinds of actual objects as learning materials, whether it be in the classroom or in the laboratory, it is important to know whether the demonstrated superiority of objects over words extends to pictures of objects as well. Furthermore, it remains to be shown whether or not differences in materials produce differences in learning efficiency when both the stimulus and the response members of paired associates are familiar.

The purpose of the present experiment thus was twofold. First, to replicate the graded facilitation associated with connective form class found by Rohwer (1966) on both printed and pictorial materials and to determine whether the increased learning efficiency produced by sentence contexts can be further augmented by the use of pictorial materials. Finally, it was of interest to determine whether or not the predicted superiority of pictorial materials would decrease with age over an interval

during which verbal facility presumably increases.

METHOD

Materials and Design

The principal factors in a $2 \times 2 \times 4$ factorial design were: grades (third vs. sixth); materials (printed vs. pictorial); and verbalization (conjunction vs. preposition vs. verb vs. control). Both the printed materials and the pictorial materials were presented by projection on a beaded screen. The pictorial study-trial materials consisted of 16-mm. black and white movie film bearing the images of 24 pairs of objects that were photographed against a neutral gray background and foreground for a total of 4 seconds each. The test-trial materials consisted of similar movie photographs of one object from each of the 24 pairs. The printed study-trial materials were pairs of words (the names of the corresponding objects in the picture materials) and the test-trial materials were the initial words from each pair photographed and mounted as 2×2 -inch slide transparencies. A complete list of the 24 pairs of objects/words is presented in Table 1.

All verbalizations were uttered by *E* as he read from a prepared script. Five-word sentences of the form article-noun-verb-article-noun were constructed for each of the 24 pairs of objects/words and constituted the materials for the verb conditions. Comparable strings for the preposition and conjunction conditions were derived from the verb strings by substituting prepositions and conjunctions, respectively, for the verbs. Verbalization for the control conditions simply consisted of *E* uttering the names of the two objects in each pair or reading the two words in each pair to *S*. A complete list of the scripts used by *E* is presented in Table 1.

Subjects

The sample consisted of 96 third- and 96 sixth-grade children drawn from a school district in a middle-class area. Two grade-level populations were sampled in order to provide information as to the relative effectiveness of picture and word materials for children of presumably lesser and greater verbal proficiency. A population younger than that in the third grade was avoided in view of the possibility that reading of the word materials might be problematical. The prediction was that the expected superiority

TABLE 1
STUDY-TRIAL MATERIALS

Strings and objects/words	Connectives
1. The FORK _____ the CAKE.	or, on, cuts
2. The TOWEL _____ the PLATE.	and, across, wipes
3. The CAT _____ the LOG.	or, over, jumps
4. The MAN _____ the POLE.	or, beside, bends
5. The BAT _____ the CUP.	and, in, strikes
6. The SHOE _____ the CHAIR.	and, beneath, taps
7. The BOAT _____ the BALL.	and, against, rolls
8. The HAND _____ the HAT.	or, inside, throws
9. The ROCK _____ the BOTTLE.	or, behind, breaks
10. The CAR _____ the WAGON.	and, under, upsets
11. The ROPE _____ the EYE.	and, around, rubs
12. The NEEDLE _____ the BALLOON.	or, outside, pops
13. The DOG _____ the GATE.	and, on, closes
14. The SPOON _____ the EGG.	and, under, rolls
15. The FIRE _____ the BED.	and, behind, burns
16. The AX _____ the WOOD.	and, upon, hits
17. The KNIFE _____ the FLOWER.	or, below, cuts
18. The BLANKET _____ the TREE.	or, around, covers
19. The MILK _____ the BOWL.	and, inside, fills
20. The TEETH _____ the APPLE.	or, near, bite
21. The HAMMER _____ the BELL.	or, over, pulls
22. The PENCIL _____ the PAPER.	or, across, tears
23. The DOLL _____ the BOOK.	and, against, opens
24. The FOOT _____ the HOUSE.	or, above, kicks

of the pictorial over the printed materials would be greater for the third- than for the sixth-grade pupils. Within each grade, 12 *Ss* were assigned randomly to each of the 8 experimental conditions.

Procedure

The task was administered individually by a study-test method for a total of two complete trials. Since two male adults served as *Es*, the assignment of *Ss* was randomized and balanced with respect to experimenters as well as with respect to experimental conditions. When an *S* entered the experimental room, he was seated in front of the projector, at a distance of approximately 10 feet from the screen. The slide and the movie projectors produced comparable levels of noise in the experimental room and the same projection screen was used with both.

The instructions informed *S* that he would be shown 24 pairs of objects/words and that he was to learn them in such a way that he could produce the name of the second member of each pair when shown the first. One example was given orally after which the first study trial commenced. During the study trial, the pairs were presented

successively at a 4-second rate, and, as each pair appeared on the screen, *E* uttered the appropriate verbalization. After a 4-second intertrial interval, the test stimuli were presented successively, again at a 4-second rate but in an order different from that of the study trials, and, as each stimulus appeared, *E* read or named it aloud. The same procedure was followed on the second trial, except that the items were presented in different orders. Responses were made orally and were recorded by *E*.

RESULTS

The amount learned was measured in terms of the total numbers of correct responses given on the two test trials. Mean numbers of correct responses are presented in Table 2 as a function of grades, materials, and verbalization. In the four-way analysis of variance performed on these data, the main effect for grades was not significant ($F = 1.41$, $df = 1/160$, $p > .05$), nor was the expected

TABLE 2
MEAN NUMBERS OF CORRECT RESPONSES AS A FUNCTION OF GRADE,
MATERIAL, AND VERBALIZATION

Material	Verbalization					
	Grade	Conjunction	Preposition	Verb	Control	Total
Printed	3	14.24	19.24	24.08	15.42	18.25
	6	13.34	23.16	22.50	19.92	19.73
Printed total		13.80	21.20	23.30	17.66	18.99
Pictorial	3	23.76	28.58	33.16	29.08	28.64
	6	30.08	25.66	33.76	27.76	29.31
Pictorial total		26.92	27.12	33.46	28.42	28.98
Total group		20.36	24.16	28.38	23.04	23.98

Note.—MSE(160) = 19.59.

interaction of Grades $\times$ Materials ($F < 1$). Clearly, however, learning was more efficient with pictorial than with printed materials ($F = 122.24$, $df = 1/160$, $p < .01$) in both grades, so much so that more than 32% of the total variance was associated with this factor.

The question whether or not the connective form class effect reported by Rohwer (1966) is replicable receives an affirmative answer in the present results. The main effect of verbalization was significant ($F = 13.63$, $df = 3/160$, $p < .01$) and, of the three string conditions, only the Verb was superior to the control, confirming the results of previous experiments in which the presentation of paired nouns in sentence contexts increased learning efficiency. As for differences associated with connective form class, more was learned with verb connectives than with prepositions which, in turn, were superior to conjunctions.² The form of the relationship between connective form class and learning efficiency, however, appears to depend upon whether the materials are words or pictures and upon the grade level of Ss. The

interaction of Materials $\times$ Verbalization was significant ($F = 2.76$, $df = 3/160$, $p < .05$), such that for printed noun pairs, the difference between the verb and preposition connectives was not significant but each was superior to conjunctions. This outcome is consistent with the results obtained by Rohwer (1966) with printed materials. In contrast, verb connectives in the pictorial conditions were superior to both prepositions and conjunctions which did not differ significantly from one another.

An examination of the three-way interaction, Grades $\times$ Materials $\times$ Verbalization ($F = 3.45$, $df = 3/160$, $p < .05$) revealed that the interaction Materials $\times$ Verbalization already described was located entirely in the sixth-grade samples ($F = 6.21$, $df = 3/160$, $p < .01$) and not in the third-grade samples ($F < 1$). For the latter Ss the effects of connective form class were virtually parallel in the printed and pictorial conditions; in both cases, the difference between verbs and prepositions was as large as the difference between prepositions and conjunctions.

The main effect of experimenters was not significant ($F < 1$) nor were any of the interactions involving this

² All post hoc comparisons were tested by means of the Scheffé method at $p = .05$.

factor, with the exception of Experimenters $\times$ Materials $\times$ Verbalization ($F = 2.91$, $df = 3/160$, $p < .05$). An appraisal of this interaction indicated that the direction of the differences obtained was consistent across *Es* but the magnitudes varied, that is, pictures were superior to words and verb connectives were associated with the greatest amount of learning for both *Es*.

In sum, the presentation of paired associates in pictorial form and in the context of sentences produces more efficient learning than any other combination of conditions examined, and the form class of context connectives is consistently associated with the amount learned, although the detailed form of that relationship appears to depend on the grade level of *Ss* and on the character of the learning materials.

DISCUSSION

With respect to the observed differences associated with type of verbalization, it is important to examine an alternative interpretation to that offered here. A careful inspection of the materials presented in Table 1 reveals that the three kinds of contextual strings differ in two ways rather than in only one. As previously noted, one of the differences is in the form class of the context connectives; the second difference lies in the number of unique words that served as connectives within each of the form classes. The set of connectives in the Verb, Preposition, and Conjunction conditions were comprised of 22, 16, and 2 unique words, respectively. If it is assumed that a contextual string connective constitutes a part of the stimulus term in each pair, then the three form-class conditions differ markedly in intralist similarity. In-

deed, the variation in intralist similarity, by itself, justifies the prediction that the amount learned should be greatest in the Verb condition, least in the Conjunction condition, and intermediate in the Preposition condition. The interpretation of the form-class effect in terms of intralist similarity, however, has already been falsified. In a previous experiment, Rohwer and Lynch (in press) systematically varied the number of different words used in each form-class condition and found that the superiority of verb strings over conjunction strings was as great when only two different verbs were used as it was when eight different verbs were used. Thus it is warranted to conclude that the differences in performance observed in the present experiment are associated with differences in the form class of the context connective and not with differences in intralist similarity.

The present demonstration of the superiority of pictorial over printed word materials for the promotion of efficient paired-associates learning raises two kinds of additional questions. The first concerns the scope, the possibility, and the manner of applying these results to the problem of presenting materials for learning in school settings. The most important restriction on the scope of application of the present results is that research to date warrants generalization only to those kinds of school learning tasks that are isomorphic with the paired-associates paradigm. Runquist and Hutt (1961), for example, report that high-school students learn verbal concepts more rapidly when the materials are represented verbally than when they are represented pictorially. The question of the possibility of applying the present results to school learning

requires an answer that is sensitive to practical as well as to scientific constraints. The most direct implication of the demonstrable superiority of the pictorial mode is that relevant curricular materials presently available in printed form should be converted to a pictorial form. Although such a conversion would not be impossible, considerable resistance might well be expected. Alternatively, it is of interest to consider the suggestion that learners themselves be trained to make covert pictorial responses to printed materials. Such a program would not only avoid the difficulties involved in reconstructing current curricular materials, it would also better equip the learner to engage in efficient acquisition regardless of the character of the content he is asked to learn. A decision as to the feasibility of such a program awaits empirical evaluation.

The second kind of question is directed at the issue of choosing an explanation for the differences observed. Data reported by Wimer and Lambert (1959) suggest that, for college-age Ss, the greater difficulty presented by the task of learning word-trigram than by that of learning object-trigram pairs is due to the greater intralist similarity among the former and not to differences in meaningfulness. Nevertheless, for Ss as young as those who participated in the present study, word and picture stimuli may differ in meaningfulness as well as in intralist similarity.

The facilitatory effect of presenting pairs in the context of sentences appears quite robust across differences in populations and differences in materials. Again, however, the task of explanation remains to be accomplished, and a clarification of the facts as to the effects of connective form class is relevant to this task. The striking differences ob-

served between the third- and sixth-grade samples in the way contextual strings affected the learning of pictorial and printed pairs invites speculation. In the third-grade samples, the relationship between connective form class and amount learned was markedly linear and consistent across materials, as if each of the kinds of learning aids provided added a constant increment to performance. For sixth-grade Ss, however, the preposition connectives appeared to be functionally equivalent to the verbs in producing superior learning of the word pairs and pictorial presentation, with only a conjunction string context, seemed sufficient to promote learning as efficient as that produced by preposition strings in the third-grade samples. Loosely speaking, it was as if the sixth-graders' threshold for engaging in facilitatory processes was lower than that of the third-graders. If this is true, it suggests again the possibility that children might be trained not only to make covert pictorial responses to printed materials but also to cast disparate learning elements in the mold of sentential structure. Regardless of the substantive worth of such speculations, they indicate the considerable amount of explanation that remains to be accomplished with regard to the present phenomena of efficient learning.

REFERENCES

- JENSEN, A. R., & ROHWER, W. D., JR. Verbal mediation in paired-associate and serial learning. *Journal of Verbal Learning and Verbal Behavior*, 1963, 1, 346-352. (a)
JENSEN, A. R., & ROHWER, W. D., JR. The effect of verbal mediation on the learning and retention of paired-associates by retarded adults. *American Journal of Mental Deficiency*, 1963, 68, 80-84. (b)
JENSEN, A. R., & ROHWER, W. D., JR. Syntactical mediation of serial and paired-

- associate learning as a function of age. *Child Development*, 1965, **36**, 601-608.
- ROHWER, W. D., JR. Constraint, syntax and meaning in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 536-540.
- ROHWER, W. D., & LYNCH, S. Semantic constraint in paired-associate learning. *Journal of Educational Psychology*, 1966, **57**, 271-278.
- ROHWER, W. D., JR., & LYNCH, S. Form class and intralist similarity in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, in press.
- ROHWER, W. D., JR., SHUELL, T. J., & LEVIN, J. R. Context effects in the initial storage and retrieval of noun pairs. *Journal of Verbal Learning and Verbal Behavior*, in press.
- RUNQUIST, W. N., & HUTT, V. H. Verbal concept learning in high school students with pictorial and verbal representation of stimuli. *Journal of Educational Psychology*, 1961, **52**, 108-111.
- WIMER, C. C., & LAMBERT, W. E. The differential effects of word and object stimuli on the learning of paired associates. *Journal of Experimental Psychology*, 1959, **57**, 31-36.

(Received September 23, 1966)

DOGMATISM AND EXAMINATION PERFORMANCE

B. JACK WHITE AND RICHARD D. ALTER

University of Utah

Whereas Ehrlich found significant negative r 's between scores on the Dogmatism Scale and scores on the true-false sociology test, Christensen found no significant r 's between dogmatism and scores on multiple-choice or essay tests in psychology. Since dogmatic Ss tend to give a disproportionate number of "true" responses on true-false tests, differences in examination formats used in the 2 studies could account for the disparity in results; however, in the present study, a comparison of correlations between dogmatism and number of correct responses to true-false items ($r = -.14$) with the correlation for multiple-choice items ($r = -.16$) produced no support for this notion. Although the disparity might in part be due to differences in course content and approach to the subject matter in sociology and psychology, the variability in r 's from 14 psychology classes in the present study indicates that sampling errors made a sizable contribution to the disparity. The weighted average r for the 14 classes was small ($-.18$; $p < .01$) and the magnitude of the r 's was variable; nevertheless, the sign of the r 's was relatively consistent (12 of 14 classes yielded negative r 's; $p = .006$).

At a common-sense level, one might expect "closed-mindedness" to interfere with performance on examinations in college courses, especially in courses where popular beliefs are likely to be contradicted. Ehrlich found support for this notion in two studies (1961a, 1961b) of the relationship between dogmatism, as measured by Rokeach's (1960) Dogmatism (D) Scale, and performance on examinations in introductory sociology. In the first study (Ehrlich, 1961a), 100 students completed the D Scale, took the same 40-item true-false examination twice during the academic quarter, then received the examination again by mail 5 months after the end of the quarter. The correlations between D scores and examination scores for the 57 students who returned the third test were $-.30$, $-.52$, and $-.54$ for the three respective tests. Ehrlich's second study (1961b) was conducted 5 years later as a follow-up. The D Scale and the examination were mailed to the 90 subjects (Ss) in the original sam-

ple who could be located and 65 of these were returned. The correlations of the resulting D scores with examination scores obtained during the course 5 years earlier were $-.32$ and $-.29$, respectively; the correlation between the most recent D scores and examination scores was $-.43$.

Despite the fact that all of the correlations were significant and in the expected direction in both studies, some additional data in the second study raises a question about the generality of the findings; namely, the correlation between D scores and overall grade averages (cumulative point-hour ratio) in college was only $-.09$. The fact that this correlation was so much smaller than the correlations between D scores and sociology examination scores led Ehrlich (1961b, p. 285) to conclude that "it seems quite likely that course *content* represents the significant source of variation."

Christensen (1963) drew similar conclusions from his study of the re-

relationship between dogmatism and grades in introductory psychology classes. The correlations between D scores and scores on a multiple-choice examination (r for 49 men = .04; r for 117 women = -.11) and between D scores and scores on an essay examination (r for men = .16; r for women = -.08) were small and insignificant—findings which led Christensen to speculate that perhaps the differences between the results of his study and Ehrlich's (1961a) were due to differences in course content and approach to the subject matter in psychology and sociology. More recently, Costin (1965) entertained the "course content" hypothesis as one possible explanation for the discrepancy between his results in introductory psychology and Ehrlich's (1961a). Costin found no significant relationship between dogmatism and multiple-choice examination scores, either on a precourse test ($N = 67$, $r = -.23$) or on a postcourse test ($N = 67$, $r = -.19$). To the writers' knowledge, the only study which has reported a significant relationship between dogmatism and grades in introductory psychology is one by Zagana and Zucher (1965) wherein a correlation of $-.20$ ($N = 517$; $p < .001$) was obtained.

One of the purposes of the present study was to investigate an alternative to the course content hypothesis mentioned above. Since dogmatic individuals evidence a response set to agree with items about which they feel uncertain (Peabody, 1961), there is a possibility that Ehrlich's (1961a, 1961b) correlations between dogmatism and sociology grades were heavily influenced by a tendency for dogmatic Ss to make a disproportionately large number of "true" responses on a true-false examination and, thus, receive poor scores on

the examination—a tendency which would not be reflected in Christensen's (1963) study since he used multiple-choice and essay examinations, or in Costin's (1965) study because he used multiple-choice items. With the foregoing in mind, assessment of the effects of examination format on the relationship between dogmatism and grades was undertaken in the present study.

The second, and more important, purpose of the present study was to obtain correlations between dogmatism and grades for a number of samples of Ss. The seven correlation coefficients reported above in connection with Ehrlich's studies (1961a, 1961b) were obtained at various times from various members of one college course, the four reported for Christensen (1963) were obtained from one sample, the two reported for Costin (1965) were based on one sample, and the one reported for Zagana and Zucher (1965) was, of course, based on one sample. Of the 14 correlations which have been mentioned, 12 were negative, a fact which hints that there is some consistency in the direction of the relationship, even if there is considerable variability in its magnitude. The present study provides additional information on the consistency of the direction and magnitude of the relationship between dogmatism and grades in a number of psychology classes taught by the same instructor and in a number of classes taught by different instructors.

METHOD

During the period 1963–1965 the D Scale was administered to 2099 students in seven instructors' introductory psychology classes at the University of Utah. Fourteen classes were involved (one instructor taught five classes, another taught four) which ranged in size from 33–319 students. Ten of the

TABLE 1
NUMBER OF SUBJECTS, MEAN D SCORES, AND CORRELATION BETWEEN
D SCORES AND EXAMINATION SCORES

Instructor	Number of subjects			Mean D score			<i>r</i> of D score & examination score		
	Class	Men	Women	Class	Men	Women	Class	Men	Women
A	75	38	37	148.3	152.5	144.0	-.28*	.07	-.42**
A	288	120	168	149.3	151.6	147.7	-.16*	-.09	-.21**
A	319	171	148	153.1	155.2	150.7	-.13*	-.20*	-.07
A	290	143	147	149.4	153.9	145.0	-.32**	-.52**	-.23**
A ^a	46	25	21	144.3	143.0	145.9	-.20	-.32	-.02
B	246	96	150	146.8	150.0	144.8	-.20**	-.34**	-.09
B	247	152	95	149.8	150.3	147.8	-.24**	-.20*	-.26*
B ^a	52	29	23	148.8	147.8	150.0	-.09	-.04	-.14
B	187	82	105	153.1	159.0	148.6	-.17*	-.29**	-.12
C	78	37	41	147.0	152.7	141.9	.03	.13	-.11
D ^a	33	21	12	140.3	147.4	128.0	-.18	-.17	-.16
E	87	58	29	144.3	145.6	141.8	.07	.07	.03
F ^a	74	39	35	146.4	149.7	128.5	-.09	-.11	-.08
G	77	45	32	152.2	156.2	146.3	-.04	-.16	.20
Totals	2099	1056	1043	149.4	152.2	146.0	-.18**	-.21**	-.15**

^a Night class.

* $p < .05$.

** $p < .01$.

classes consisted of regular daytime students, most of whom were freshmen and sophomores. The four other classes were taught at night and were comprised of both working adults and daytime students in a ratio of approximately 2:1.

One instructor (referred to later as Instructor "B") used both true-false and multiple-choice questions in his examinations whereas all other instructors used only multiple-choice questions. The text and the examination questions which were used were the option of the instructor; however, all of the texts were of approximately equal caliber and most of the questions came from the instructor's manual provided with the text.

The D Scales were administered during a class period 3-5 weeks before the end of the quarter. At the end of the quarter the total number of questions correctly answered by each student on the various examinations was obtained and the resulting total scores were correlated with D scores.

RESULTS

As can be seen in Table 1, the statistically significant correlations between D scores and examination

scores tend to come from the larger classes, as would be expected. Although the correlations for the 14 classes are generally small (weighted average $r = -.18$; $p < .01$), there is considerable consistency in the sign of the correlations. According to binomial tests, the fact that 12 of the 14 classes yielded negative correlations is significant ($p = .006$), as is the fact that women in 12 of the 14 classes yielded negative correlations ($p = .006$) and the fact that men in 11 of the classes did likewise ($p = .029$).

The correlations in Table 1 show a good deal of variability in the relationships between D scores and examination scores, both between classes taught by the same instructor (where course content and approach are likely to be more or less constant) and between classes taught by different instructors (where, of course, both course content and approach are

likely to vary). While the range of correlations for the 14 classes was not especially large ($-.32$ to $.07$), the range for men ($-.52$ to $.13$) and the range for women ($-.42$ to $.20$) was rather large. In a speculative vein, it is interesting to note that if the men and the women who produced the largest negative correlations in Table 1 had happened to be in the same class, the weighted average correlation would have been $-.50$. Had similar circumstances obtained for the men and women who produced the largest positive correlations in Table 1, the weighted average correlation would have been $.16$. Under these hypothetical circumstances, the resulting correlations would have approximated the range of the most disparate of the 11 correlations reported by Ehrlich (1961a) and Christensen (1963); namely, $-.54$ (Ehrlich) to $.16$ (Christensen).

Dogmatism, Examination Format, and Examination Scores

As indicated earlier, Instructor B in Table 1 used both true-false and multiple-choice questions on each of his examinations. During the academic quarter, the 187 students in the fourth class listed for Instructor B in Table 1 took five examinations which contained a total of 347 questions, of which 118 (34%) were true-false and 229 (66%) were multiple-choice. The correct answer for 44% of the true-false items was "true."

The correlation between D scores and the number of questions answered with "true" responses was $.23$ ($p < .01$). Despite this, performance on the true-false items was no worse than performance on the multiple-choice items—the correlation between D scores and number of correct responses to the true-false items was

$-.14$ and the correlation for the multiple-choice items was $-.16$. Thus, the fact that Ehrlich used true-false items does not appear to account for the differences between his results and those of Christensen or Costin.

DISCUSSION

The results of the present study indicate that it is unlikely that the differences between Ehrlich's (1961a, 1961b) results and those of Christensen (1963) or Costin (1965) were due to differences in examination format. While variations in course content may account for some proportion of the differences among the studies just mentioned, the variability in the correlations obtained in the present study suggests that sampling errors also made a sizable contribution to the differences.

Despite the fact that the correlations between D scores and examination scores were rather consistently negative in the present study, the weighted average correlation was small and the variability in the magnitude of the correlations was rather large. Thus, it seems fair to say that the predictive power of the D Scale with regard to grades is not impressive.

REFERENCES

- CHRISTENSEN, C. M. A note on "Dogmatism and Learning." *Journal of Abnormal and Social Psychology*, 1963, **66**, 75-76.
- COSTIN, F. Dogmatism and learning: A follow-up of contradictory findings. *Journal of Educational Research*, 1965, **59**, 186-188.
- EHRLICH, H. J. Dogmatism and learning. *Journal of Abnormal and Social Psychology*, 1961, **62**, 148-149. (a)
- EHRLICH, H. J. Dogmatism and learning: A five-year follow-up. *Psychological Reports*, 1961, **9**, 283-286. (b)
- PEABODY, D. Attitude content and agreement

set in scales of Authoritarianism, Dogmatism, Anti-Semitism, and Economic Conservatism. *Journal of Abnormal and Social Psychology*, 1961, **63**, 1-11.

ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.

ZAGONA, S. V., & ZUCHER, L. A., JR. The relationship of verbal ability and other cognitive variables to the open-closed cognitive dimension. *Journal of Psychology*, 1965, **60**, 213-219.

(Received October 11, 1966)

A TWENTY-COLLEGE STUDY OF STUDENT $\times$ COLLEGE INTERACTION USING TAPE (TRANSACTIONAL ANALYSIS OF PERSONALITY AND ENVIRONMENT): RATIONALE, RELIABILITY, AND VALIDITY¹

LAWRENCE A. PERVIN

Princeton University

A study of college characteristics and Student $\times$ College interaction using Transactional Analysis of Personality and Environment (TAPE), an instrument based on the semantic differential. 3,016 students from 21 colleges rated the following concepts on the 52 scales in Form A or the 52 scales in Form B: My College, My Self, Students, Faculty, Administration, Ideal College. Ratings of satisfaction with aspects of college life were also made on 16 scales. Data relevant to 4 areas are presented: (a) TAPE as a measuring device for intra- and interinstitutional research; (b) the relationship between concept discrepancy scores and satisfaction ratings; (c) reliability of TAPE; (d) factorial structure of TAPE. In general, discrepancies between student perceptions of themselves and their college were found to be related to dissatisfaction with college. The data were interpreted as supporting the theoretical model of Student $\times$ College interaction and the utility of TAPE in this area of research. This research should be useful in suggesting the transactions within the college, or between students and parts of the college, that might be influenced in the direction of fostering student development.

Social science research has often attempted to account for behavior in terms of the individual or the environment. This dichotomy is seen in controversies between sociology and psychology, social psychology and clinical psychology, S-R theory and psychoanalysis, and in the nature-nurture controversy (Hunt, 1965). The theoretical rationale for the research reported here and for the development of the Transactional Analysis of Personality and Environment (TAPE) questionnaire is that human behavior can be best understood in terms of the interactions or transactions (Dewey & Bentley, 1949) between the individual and his environment. This view has been most

clearly expressed in the need-pressure personality theory of Murray (1938) though it has also received attention in, among others, the theoretical systems of Lewin (1951) and Heider (1958).

The interaction approach has been found to be useful in areas outside the academic setting such as interpersonal attraction (Lott & Lott, 1965; Miller, 1963; Newcomb, 1956), occupational choice and satisfaction (Super, 1963), adaptation to cultural patterns (Jahoda, 1961), and psychopathology (French, 1963; Kelly, 1966; Spiegel, 1957). Within the academic setting, performance has been related to an interaction between student personality and demands of the curriculum (Malleon, 1959; Snyder, 1966), situational stress (Smith & Rockett, 1958), instructor personality (McKeachie, 1961), and type of exam (Claunch, 1964). Research on col-

¹This research was supported by grants from the Straus Council on Human Relations, the National Institute of Mental Health (MH-08321), and the United States Office of Education.

lege dropouts suggests the importance of a match or fit between student and college (Pervin & Rubin, in press; Stern, 1962; Summerskill, 1962).

While instruments have been developed to measure the college environment (Astin, 1963; Pace, 1963; Stern, 1963), they do not provide for an analysis in terms of Individual $\times$ Environment interaction. The Activities Index (AI) and College Characteristics Index (CCI) were developed to follow Murray's need-press system, but they have not been used in this way and it is not clear that the need and press scales on the instruments are comparable. Also, while the AI and CCI include items relevant to various parts of a college environment (students, faculty, administration), analyses in terms of the interactions or transactions among these parts are not generally reported.

As Hunt (1965) notes, the semantic differential represents "an important method of assessing the interaction between people and situations [p. 83]." The research reported here describes the current status of TAPE, an instrument which uses the semantic differential technique to study the various interactions and transactions that occur within a college environment, and their relevance to institutional strain and student satisfaction. In sum, this report discusses TAPE data in relation to the three kinds of analysis for which it was developed: (a) comparisons of different college environments; (b) analysis of sources of conflict or strain within a college environment and comparisons of these sources across colleges; and (c) the analysis of individual performance and satisfaction as a function of Student $\times$ College interaction. "The organism which adapts well under one condition would not

survive under another. If for each environment there is a best organism, for every organism there is a best environment [Cronbach, 1957, p. 679]."

METHOD

Subjects

Participants in this study were 3,016 students from 21 colleges.^{*} For Form A of TAPE, there were 1,393 subjects (Ss) from 11 public and 10 private colleges while for Form B there were 1,623 Ss from 11 public and 9 private colleges. The colleges in the sample varied in geographical location, size, and male-female ratio. Also, there was an attempt to select colleges with different campus atmospheres.

The selection of Ss differed at the various participating colleges. Generally the sample consisted of students taking an introductory psychology course. In one case volunteers were used and in a few there was a random sampling of students in each of the undergraduate years. In all cases the students were not paid and, except for two colleges, participation was anonymous. Data were collected in the spring of 1966.

Materials and Procedure

TAPE is based on the semantic differential technique and asks students to rate a number of concepts on the same polar adjective scales. In its standard form TAPE requires that the following concepts be judged on 52 scales: College, Self, Students, Faculty, Administration, Ideal College. Concepts relevant to the college refer to the one the student is attending. In this study 11-point scales were used as opposed to the more traditional 7-point scales. This followed Gulliksen's (1958) suggestion that Ss are capable of making finer discriminations than are generally allowed for on the semantic differential.

Forms A and B follow the same format but contain different scales. The scales used in these forms were developed on an

^{*}Participating colleges and universities were: Antioch, Brooklyn, Bucknell, Cincinnati, Dartmouth, Douglass, Georgetown, Haverford, Kent, Maryland, Middlebury, New Mexico, North Carolina, Pennsylvania, Princeton, Smith, South Dakota, Stony Brook, Tennessee, Texas, Wesleyan.

a priori basis as to which dimensions might be important in assessing Individual $\times$ Environment interaction, particularly Student $\times$ College interaction. Modifications were then made on the basis of a 10-college exploratory study. Scales were chosen on the basis of whether they discriminated among colleges and between concepts; that is, whether a distribution of responses was obtained across subjects (colleges) or concepts. Examples of the polar adjective scales are: authoritarian-democratic, grinding-fun-loving, religious-secular, idealistic-materialistic, equalitarian-status-oriented.

The TAPE questionnaire included a beginning page on which students gave information such as college class, sex, area of concentration and whether a resident or a commuter, and also included 16 questions relating to satisfaction with the college environment. The latter appeared in the middle of the questionnaire between the Students and Faculty concept ratings. Ratings were made on 11-point scales with the extremes being defined for Ss. About 45 minutes was required for a student to fill in the background information, rate each of the six concepts on 52 scales, and complete the 16 satisfaction items.

RESULTS

The results are presented in relation to the three purposes for which the instrument was developed: intra-institutional research, interinstitutional research and Student $\times$ College interaction. Data relevant to the first two goals are largely descriptive, whereas the data in relation to the third are more easily interpreted in terms of statistical tests of significance. Data relevant to the reliability and factor structure of TAPE are also presented.

Intra- and Interinstitutional Comparisons: Descriptive Properties of TAPE

Initial intrainstitutional analysis of TAPE data consists of looking at mean scale ratings on each concept and comparisons of means for a scale across concepts; that is, how do students at a college perceive them-

selves and different parts of their college environment? Interinstitutional comparisons include those of scale means for different colleges on single concepts and of the relative distribution of means for a single scale across different concepts; that is, do students at various colleges perceive different characteristics as associated with individual parts of the college environment (students, faculty, administration, etc.) and in the pattern of characteristics associated with these parts? Examples of these kinds of comparisons are given in Figure 1, where mean ratings on the concepts are plotted for three colleges on each of two scales.

The data presented in Figure 1 illustrate a number of conclusions relevant to TAPE: (a) There is considerable variability in scale ratings across concepts for a single college. For example, on the first scale illustrated it can be seen that for College 1 the college and administration are rated as quite conservative, the self and students less so, and the ideal college as quite liberal. (b) There is considerable variability across colleges in scale ratings on each concept and, perhaps more importantly, in the pattern of ratings for a scale across the six concepts. For example, in the second scale in Figure 1 the means for Schools 1 and 2 are quite similar except for the student concept. The absolute values for Schools 2 and 3 are quite different but, with the exception of ideal college, the pattern of means is quite similar. On some scales almost all of the schools show a similar pattern whereas on other scales almost every possible pattern occurs. (c) The variability in pattern of mean ratings indicates that there can exist large discrepancies between the way students see various parts of the college.

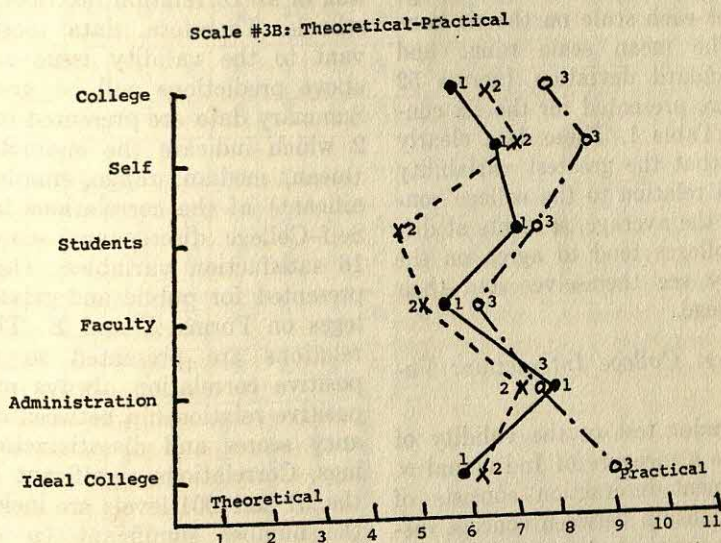
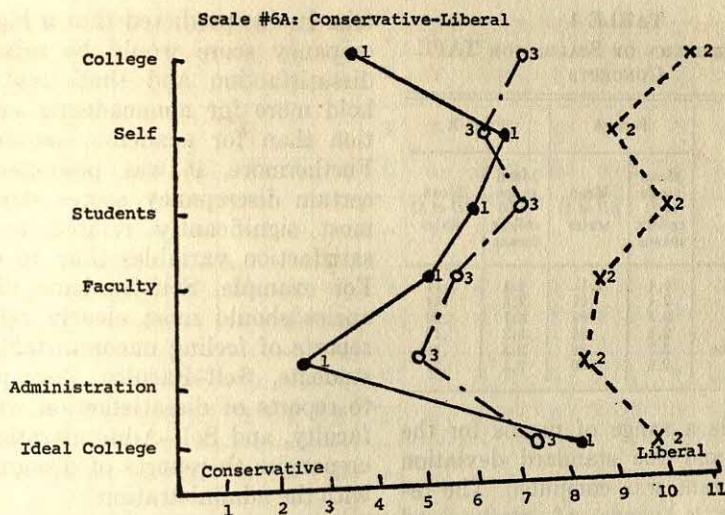


FIG. 1. Concept means for three colleges on two TAPE scales.

Particularly noteworthy is the fact that the way students see themselves can be at variance with the way they see students at the college in general (School 2, Scale 3B). It can be suggested that large discrepancies suggest sources of strain in the functioning of the system. Different scales and concepts are relevant to different colleges in relation to the strain is-

sue. (d) The mean rating for college can be more extreme than the mean ratings for the parts of the college; the whole can be greater than or different from the sum of its parts (College 2, Scale 6A).

As an indication of the variability in ratings across colleges, means for the colleges were computed for each scale on the six concepts. Then, for

TABLE 1
CHARACTERISTICS OF SCALES FOR TAPE
CONCEPTS

Concept	Form A		Form B	
	Mean range of 21 college means	Mean SD of 52 scales	Mean range of 20 college means	Mean SD of 52 scales
College	4.4	1.17	4.4	1.11
Self	2.4	.62	2.5	.64
Students	3.5	.86	3.7	.89
Faculty	2.5	.65	3.2	.78
Administration	3.2	.80	3.3	.80
Ideal College	2.6	.60	3.1	.77

each scale a range of means for the colleges and the standard deviation of the means was computed. The result was a range of means and standard deviations for the college means for each scale on the six concepts. The mean scale range and scale standard deviation (across 52 scales) are presented for the six concepts in Table 1. These data clearly indicate that the greatest variability occurs in relation to the college concept. On the average, students at different colleges tend to agree on the way they see themselves and their ideal college.

Student $\times$ College Interaction: Validity

The major test of the validity of TAPE as a measure of Individual $\times$ Environment interaction consists of the relationship between concept discrepancy scores and satisfaction ratings. For each *S* a discrepancy score was calculated for each pair of concept ratings ($N = 15$). A discrepancy score represented the sum of the absolute difference in ratings of two concepts on 52 scales. For each TAPE form correlations were computed between the concept discrepancy scores and the satisfaction ratings completed in the middle of the questionnaire—a 15×16 correlation ma-

trix. It was predicted that a high discrepancy score would be related to dissatisfaction and that this would hold more for nonacademic satisfaction than for academic satisfaction. Furthermore, it was predicted that certain discrepancy scores should be most significantly related to some satisfaction variables than to others. For example, Self-Students discrepancies should most clearly relate to reports of feeling uncomfortable with students, Self-Faculty discrepancies to reports of dissatisfaction with the faculty, and Self-Administration discrepancies to reports of disagreement with the administration.

Space does not permit the presentation of all correlation matrices for all schools. Therefore, data most relevant to the validity issue and the above predictions will be presented. Summary data are presented in Table 2 which indicate the characteristics (mean, median, range, number significant) of the correlations between Self-College discrepancy scores and 16 satisfaction variables. These are presented for public and private colleges on Forms A and B. The correlations are presented so that a positive correlation always means a positive relationship between discrepancy scores and dissatisfaction ratings. Correlations significant beyond the .01 and .001 levels are included in the number significant ($p < .05$) category. The maximum number of significant correlations would be the same as the number of colleges in the sample.

The data in Table 2 give clear support to the predictions. While the range of correlations indicates some variability in the stability of the relationships across schools and items, the general trend is clearly in the direction of a relationship between high Self-College discrepancy and dis-

TABLE 2

CORRELATIONS BETWEEN SELF-COLLEGE DISCREPANCY SCORES AND SATISFACTION RATINGS FOR PUBLIC AND PRIVATE COLLEGES

	TAPE Form A										TAPE Form B									
	Public					Private					Public					Private				
	M	Mdn	Range	No. significant r^a	M	Mdn	Range	No. significant r^a	M	Mdn	Range	No. significant r^a	M	Mdn	Range	No. significant r^a	M	Mdn	Range	No. significant r^a
1. How likely is it that you will at some time drop out of college?	.13	.16	-.36/.69	2	.21	.28	-.10/.42	6	.15	.14	.06/.43	1	.23	.25	-.05/.56	6				
2. How likely is it that you will at some time drop out for academic reasons?	-.02	.03	-.33/.20	0	.08	.09	-.12/.22	2	.07	.09	-.11/.13	0	.06	.08	-.06/.15	0				
3. How likely is it that you will drop out for nonacademic reasons?	.19	.16	-.02/.76	2	.29	.33	-.04/.46	7	.16	.16	.01/.35	4	.30	.33	-.07/.59	7				
4. How often do you think about dropping out for nonacademic reasons?	.23	.18	.05/.56	3	.32	.35	-.01/.43	8	.30	.29	.05/.71	7	.36	.37	.11/.66	8				
5. How comfortable do you feel with most of the students at your college?	.21	.13	-.06/.69	5	.29	.37	.03/.49	6	.25	.27	.01/.65	8	.31	.33	.07/.46	7				
6. How similar do you feel your values are to the values of the faculty?	.17	.18	-.15/.53	5	.30	.30	.14/.41	9	.25	.32	-.42/.46	9	.30	.24	.06/.55	6				
7. How much do you agree with the administrative rules and regulations?	.26	.29	-.01/.45	8	.30	.30	.06/.55	8	.32	.29	.00/.70	9	.37	.38	.05/.60	7				
8. How much do you disagree with your college on important issues?	.17	.14	-.08/.46	3	.23	.21	.09/.44	6	.27	.23	-.03/.74	6	.29	.27	.18/.46	6				
9. How often do you feel out of place at your college?	.26	.16	.09/.77	4	.39	.38	.17/.50	9	.29	.30	.15/.41	7	.38	.42	.13/.51	7				
10. In terms of your own needs and desires, how satisfied are you with the academic aspects of your college?	.18	.12	-.18/.63	5	.26	.30	-.14/.58	6	.29	.27	-.03/.65	8	.29	.31	.00/.60	6				
11. In terms of your own needs and desires, how satisfied are you with the nonacademic aspects of your college?	.35	.37	.03/.72	9	.35	.32	.20/.54	10	.22	.24	-.08/.38	7	.40	.39	.20/.60	7				
12. So far, what kinds of times have you had at your college?	.24	.18	.06/.50	6	.28	.31	.10/.58	8	.25	.22	.11/.68	7	.31	.34	-.11/.47	7				
13. Do you think that your academic experience would have been more rewarding elsewhere?	.22	.26	-.07/.50	5	.37	.42	.00/.57	8	.33	.31	.16/.57	8	.38	.42	-.10/.58	8				
14. Do you think that your nonacademic experience would have been more rewarding elsewhere?	.24	.19	-.09/.59	6	.26	.25	.05/.49	8	.27	.23	.05/.53	7	.27	.25	.16/.39	5				
15. To what extent is the college responsible for your frustrations in relation to academic goals?	.19	.16	-.03/.48	3	.16	.12	.04/.39	1	.18	.25	-.02/.31	5	.17	.14	.06/.50	2				
16. To what extent is the college responsible for your frustrations in relation to nonacademic goals?	.17	.13	-.07/.64	2	.15	.13	-.14/.50	3	.11	.11	-.09/.33	2	.18	.25	-.12/.57	5				

Note.—For TAPE Form A, $N = 11$ public, 10 private; for TAPE Form B, $N = 11$ public, 9 private.^a Correlations significant at $p < .05$, 1-tailed test.

TABLE 3

MEAN AND MEDIAN CORRELATIONS BETWEEN THREE DISCREPANCY SCORES AND THREE SATISFACTION VARIABLES FOR PUBLIC AND PRIVATE COLLEGES

Concept discrepancy	Question number	Tape Form A				Tape Form B			
		Public		Private		Public		Private	
		<i>M</i>	<i>Mdn</i>	<i>M</i>	<i>Mdn</i>	<i>M</i>	<i>Mdn</i>	<i>M</i>	<i>Mdn</i>
Self-Students	5	.30	.32	.37	.39	.32	.29	.39	.40
	6	.18	.21	.18	.17	.25	.28	.29	.30
	7	.19	.17	.19	.21	.27	.23	.31	.38
Self-Faculty	5	.14	.09	.17	.19	.19	.16	.23	.24
	6	.23	.30	.35	.37	.33	.38	.44	.47
	7	.18	.26	.27	.27	.25	.24	.32	.34
Self-Administration	5	.08	.06	.19	.18	.15	.16	.20	.26
	6	.08	.11	.22	.24	.30	.35	.30	.28
	7	.31	.38	.44	.42	.41	.36	.49	.49

Note.—For TAPE Form A, $N = 11$ public, 10 private; for TAPE Form B, $N = 11$ public, 9 private.

satisfaction. The relationships are relatively stable across the two forms, though there is some tendency for the public college correlations to be higher on Form B than on Form A. Since dropping out of college is a much more complex phenomenon than satisfaction, it is not surprising that correlations for the former are generally lower than those for the latter. Though there are some exceptions, the correlations relating to academic dropout and satisfaction are generally lower than those for non-academic dropout and satisfaction.

While not presented here, the correlations between College-Ideal College discrepancy scores and satisfaction ratings are very similar to those for Self-College discrepancy scores and satisfaction, though generally they are slightly higher. It was predicted that certain discrepancy scores should correlate better with some satisfaction variables than with others and data relevant to this prediction are presented in Table 3. Satisfaction Questions 5, 6 and 7 related

to satisfaction with students, faculty, and administration respectively. We would expect correlations between Self-Students, Self-Faculty, and Self-Administration discrepancies to correlate best with the corresponding satisfaction variables. The data in Table 3 indicate that while the discrepancy scores generally relate to a variety of satisfaction items, in every case the discrepancy score correlates highest with the corresponding satisfaction variable. This holds for public and private colleges on both forms.

Reliability

The test-retest reliability of the semantic differential has been found to be quite high (Miron, 1961; Os-good, Suci, & Tannenbaum, 1957, pp. 126-143). Thus, in part the reliability of TAPE stands upon past research on the semantic differential. However, some data can be presented which are specifically relevant to the reliability of TAPE as a measuring instrument. A reliability study was

completed on Form B by James Pedersen at South Dakota State University. In the spring semester students completed TAPE Form B and then a sample of these completed ratings on three of the concepts about a month later. All Ss rated the Self concept while one-half of the sample also rated the College concept and the other half of the sample the Students concept.

Data relevant to the reliability of TAPE Form B are presented in Table 4. First the reliability of individuals can be looked at. The first column indicates the mean reliability of ratings for individual Ss on each of the three concepts. These were computed across 52 scales for each S rating a concept. The data are quite comparable to those reported by Lilly (1965), which represents the one other place this kind of reliability is reported for the semantic differential. There was considerable variation among individuals in their reliability correlations. Another reliability check for individuals is that on discrepancy scores. If the individual had a large discrepancy score between a pair of concepts on one occasion did he also have a large discrepancy score upon the second occasion? Product-moment correlation coefficients were run for Self-College and Self-Student discrepancy scores, obtained for the same Ss from the two testing sessions. The correlation for the former was .87 and that for the latter .95. The conclusion can be drawn that the discrepancy scores show a quite high degree of reliability. Even if an individual tends to vary in his scale ratings, in general he tends to see about the same amount of discrepancy between two concepts when the time between the ratings is not very great.

Turning to the scale ratings, the

TABLE 4
PRODUCT-MOMENT CORRELATION
COEFFICIENTS INDICATING
RELIABILITY OF TAPE
FORM B

Type of reliability	Concept		
	College	Self	Students
1. Mean of individual subject reliabilities	.59	.70	.58
2. Mean of 52 scale reliabilities	.40	.56	.47
3. Concept reliability-across scales + Ss	.58	.70	.60
4. Scale reliability-means for two samples	.98	.98	.98
5. Scale reliability-means for test-retest	.95	.99	.95

Note.—For College, $N = 37$; for Self, $N = 75$; for Students, $N = 35$.

second column in Table 4 indicates the correlation between two sets of ratings on the scales. Here the ratings for each scale are correlated across Ss, resulting in 52 correlation coefficients for each concept. The correlations here are not high but are similar to those reported by Norman (1959). Actually, these correlations do not represent a fair appraisal of the scale reliability. For example, the first scale on the college concept had a product-moment correlation coefficient of .14—quite low. Yet, 50% of the ratings were either the same or one position off and 75% of the ratings were two or fewer positions off. The problem is that if a scale has low variability in ratings, often a desirable characteristic, and some minor amount of changes in ratings, the reliability estimate will tend to be low.

Another kind of reliability is that of concept reliability. As a measure of this the ratings for the two occasions were correlated across scales and Ss. These correlations range from .58 to .70. As a more global measure of the reliability of the ratings, the ratings from the two occasions were

TABLE 5
SCALE FACTORS AND SAMPLE SCALES
DERIVED FROM THREE-MODE
FACTOR ANALYSIS

Factor	Sample scales
1. Impulsivity-Inhibition	1. sober-intoxicated disciplined-undisciplined
2. Humane idealism-Narcissism	2. humane-self-interested idealistic-materialistic
3. Warm-Cold	3. warm-cold sociable-unsociable
4. Introversion-Extroversion	4. introverted-extroverted eggheadish-well-rounded
5. Goal-directed activity	5. motivated-undirected industrious-tranquil
6. Liberal idealism-Conservative pragmatism	6. social welfare-laissez faire socialistic-capitalistic idealistic-materialistic
7. Scholarship	7. research-application scholarly-nonscholarly
8. Optimism-Alienation	8. relaxed-tense optimistic-pessimistic
9. Conventionality	9. religious-secular moral-amoral
10. Creativity	10. artistic-pragmatic esthetic-task-oriented
11. Sensitivity	11. feminine-masculine sensitive-insensitive
12. Tradition	12. upperclass-middle class elegant-common traditional-traditionless
13. Cosmopolitan-Provincial	13. cosmopolitan-provincial urban-rural

correlated across concepts, scales, and Ss ($N = 7,644$). The result was a product-moment correlation coefficient of .65.

Finally, there is the reliability of the scale means. The scale means for a large sample were correlated with those of a smaller sample for each of the six concepts. The data for three concepts are given in the fourth column of Table 4. The correlation coefficients for the other three concepts and for the satisfaction items were either .98 or .99, indicating a high degree of reliability. Correlations between the scale means for the first time TAPE was administered and the scale means for the second testing session (test-retest reliability) are in-

dicated in the last column in Table 4. These data are comparable to those reported by Jenkins, Russell, and Suci (1958) and clearly indicate that the scale means are quite reliable between samples and testing sessions.

Factor Structure

A three-mode factor analysis was completed on the TAPE data across both forms.³ The procedure followed was the same as that in Levin's (1965) analysis, which was based on work by Tucker (1964). Whereas the usual semantic differential analysis examines only the scale-mode factors, the three-mode factor analysis permits the investigator to explore the factors in the scale, concept, and subject (college) modes simultaneously as well as the interrelations among these three sets of factors.

In this study the factor analysis was completed across 104 scales, 6 concepts, and 20 colleges. College means were used to represent the subject mode. Fourteen scale factors, three concept factors, and three college factors were derived. All concepts were important in the concept factors and the college factors suggested one for state colleges and two for private colleges. The latter seemed to consist of a group of elite conservative colleges and a group of elite liberal and less conventional colleges. The addition of other colleges in the future will likely increase the number of college factors.

The scale factors are given in Table 5 along with two examples of scales with high loadings on the factors. The factors clearly cover a variety of areas including those of modes of impulse expression, interests or goals, and value orientations. A number of

³ The three-mode factor analysis was completed by Roy S. Lilly.

factors appear to be similar to those reported by Pace (1963) in the development of CUES.

DISCUSSION

The data in this study have been presented in relation to the specific goals of the research. However, the data in relation to each goal have relevance for one another, for the theoretical rationale involved, and for the TAPE instrument as a whole. TAPE appears to hold considerable promise for intra- and interinstitutional research. The validity data suggest that discrepancy scores can be useful in institutional research. For example, mean discrepancy scores can be computed for each pair of concepts. Comparisons of these would suggest sources of strain in the functioning of the parts of the college or differences in the system functioning of different colleges. In fact, large mean discrepancy scores for colleges have tended to be related to large mean dissatisfaction scores, a relationship already noted in relation to individuals.

The discrepancy score-satisfaction data can also be used to assess the scales or dimensions upon which the greatest discrepancies occur and those which are most related to dissatisfaction. An analysis of two schools along these lines suggests that significant areas of discrepancy and dissatisfaction vary from college to college and the relationship between perceptions may vary within the same area. For example, dissatisfied students at one school saw the college as more conservative, less equalitarian, and less scholarly, and the self as more liberal, more equalitarian, and more scholarly than did satisfied students. On each of these scales the relationship was reversed for dissatisfied students at the sec-

ond school; that is, at the second school dissatisfied students saw the college as more liberal, more equalitarian, and more scholarly, and the self as less liberal, less equalitarian, and far less scholarly than did satisfied students. Indeed, data such as these strongly support the conclusion of Douvan and Kaye (1962) that there is something wrong in the process by which students select colleges and that the time may come when we are able to arrive at a student-college fit which is most conducive to developmental growth and change. These data do not suggest that homogeneity of colleges or homogeneity of students within a college is best. Rather they suggest that there is an optimum fit between student and college, the qualities of which will vary for different students and different colleges. Viewed in this light, this research should be useful in suggesting the transactions within the college, or between students and parts of the college, that might be influenced in the direction of fostering student development.

Cronbach (1958) has suggested that discrepancy scores may be less useful than using ratings of one or another concept alone. An intensive analysis of the data from one school suggested that this was not the case for the relationships reported here. Discrepancy scores appeared to account for more of the variance than either self or college ratings. Furthermore, since the relationship between particular scale ratings and satisfaction scores can vary from college to college, relationships involving discrepancy scores will likely have greater stability across schools.

Another possible source of variance in the data investigated was that of a curvilinear relationship between size of discrepancy score and degree

of satisfaction. The data for a number of the correlations were plotted but evidence for such a curvilinear relationship was not found.

It is possible to argue that the satisfaction ratings were contaminated by the ratings of the concepts. For this to be true, one would have to argue that ratings on the first three concepts biased the satisfaction ratings and the latter in turn biased ratings on the next three concepts. If this were the case, one would suspect a general bias in ratings across satisfaction items. Yet, discrepancy score-satisfaction relationships varied between academic and nonacademic kinds of satisfaction and depended upon the concept pair and satisfaction item involved. Specific discrepancy scores (Self-Students, Self-Faculty, Self-Administration) tended to have their highest correlations with the corresponding satisfaction items. Finally, in an earlier study (Pervin & Rubin, in press) similar relationships were found even though concept and satisfaction ratings were made a week apart.

Results from the three-mode factor analysis should be useful in future research with TAPE. Data can now be analyzed in terms of how each of the concepts for each college loads on the scale factors. This provides intra- and interinstitutional comparisons using factor scores rather than scale scores. Furthermore, the development of factors now allows for an analysis of the relationship between semantic space discrepancy scores (Osgood et al., 1957) and satisfaction ratings. Also, the suggestion of Cronbach (1958) that the factor scores may vary in their relationships to the dependent variable can be investigated.

Future research with TAPE can follow along a number of lines. Four

areas can be specified: (a) intra-institutional variables affecting correlations. Here one may study differences between males and females, members of different colleges, or members of different college years. Some early analyses suggest that the relationships hypothesized hold best for the freshman year. (b) Inter-institutional variables affecting correlations. Here one may study the characteristics (size, complexity, etc.) of different colleges which affect the nature of the relationships. It has already been suggested that some scales and factors may be more important for some institutions than for others. Also, some colleges may show greater tolerance for diversity and heterogeneity. (c) Personality variables. It may be that some individuals are more tolerant of differences and are more flexible in adapting to them than others. (d) Instrument variables. Analysis of individual scale and factor scores has already been suggested. Current research is also being directed toward an analysis of the direction of perceived discrepancies rather than just analyzing distance. A hypothesis being investigated here is that discrepancies which are perceived as helping the individual become like his ideal self are desirable whereas those which are perceived as taking him away from the ideal self are not. Other studies involving TAPE include the comparison of faculty ratings with student ratings and the analysis of changes in student perceptions over time.

In contrast to other instruments, TAPE allows for the analysis of transactions among parts of the college system and uses student perceptions of areas of interest as opposed to defining the characteristics relevant to the area for the student (Astin, 1965). The data presented are

taken as supporting the theoretical rationale which led to the development of TAPE and the utility of TAPE in the study of Student $\times$ College interaction.

REFERENCES

- ASTIN, A. W. Further validation of the environmental assessment technique. *Journal of Educational Psychology*, 1963, **54**, 217-226.
- ASTIN, A. W. The inventory of college activities: Assessing the college environment through observable events. Paper presented at the meeting of the American Psychological Association, Chicago, September 1965.
- CLAUNCH, N. Cognitive and motivational characteristics associated with concrete and abstract levels of conceptual complexity. Unpublished doctoral dissertation, Princeton University, 1964.
- CRONBACH, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1957, **12**, 671-684.
- CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford University Press, 1958. Pp. 353-379.
- DEWEY, J., & BENTLEY, A. F. *Knowing and the known*. Boston: Beacon, 1949.
- DOUVAN, E. & KAYE, C. Motivational factors in college entrance. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1962. Pp. 193-224.
- FRENCH, J. R. P. The social environment and mental health. *Journal of Social Issues*, 1963, **19**, 39-56.
- GULLIKSEN, H. How to make meaning more meaningful. *Contemporary Psychology*, 1958, **3**, 115-119.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- HUNT, J. McV. Traditional personality theory in the light of recent evidence. *American Scientist*, 1965, **53**, 80-96.
- JAHODA, M. A social-psychological approach to the study of culture. *Human Relations*, 1961, **14**, 23-30.
- JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *American Journal of Psychology*, 1958, **71**, 688-699.
- KELLY, J. G. Ecological constraints on mental health services. *American Psychologist*, 1966, **21**, 535-539.
- LEVIN, J. Three-mode factor analysis. *Psychological Bulletin*, 1965, **64**, 442-452.
- LEWIN, K. *Field theory in social science*. New York: Harper & Row, 1951.
- LILLY, R. A developmental study of the semantic differential. Unpublished doctoral dissertation, Princeton University, 1965.
- LOTT, A. J., & LOTT, B. E. Group cohesiveness as interpersonal attraction. *Psychological Bulletin*, 1965, **64**, 259-309.
- MALESON, N. Operational research in the university. *British Medical Journal*, 1959, **1**, 1031-1035.
- MCKEACHIE, W. J. Motivation, teaching methods, and college learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1961*. Lincoln: University of Nebraska Press, 1961. Pp. 111-142.
- MILLER, D. R. The study of social relationships: Situations, identity, and social interaction. In S. Koch (Ed.), *Psychology: A study of a science*, Vol. 5. New York: McGraw-Hill, 1963. Pp. 639-737.
- MIRON, M. S. The influence of instruction modification upon test-retest reliabilities of the semantic differential. *Educational and Psychological Measurement*, 1961, **21**, 883-893.
- MURRAY, H. A. *Explorations in personality*. New York: Oxford University Press, 1938.
- NEWCOMB, T. M. The prediction of interpersonal attraction. *American Psychologist*, 1956, **11**, 575-586.
- NORMAN, W. T. Stability characteristics of the semantic differential. *American Journal of Psychology*, 1959, **72**, 581-584.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- PACE, C. R. CUES: Preliminary manual. Princeton: Educational Testing Service, 1963.
- PERVIN, L. A., & RUBIN, D. B. Student dissatisfaction with college and the college dropout: A transactional approach. *Journal of Social Psychology*, in press.
- SMITH, W. F., & ROCKETT, F. C. Test performance as a function of anxiety, instructor and instructions. *Journal of Educational Research*, 1958, **52**, 138-141.
- SNYDER, B. R. Adaptation, education, and emotional growth. In L. A. Pervin, L. E. Reik, & W. Darymple (Eds.), *The college dropout and the utilization of talent*. Princeton: Princeton University Press, 1966. Pp. 155-175.
- SPIEGEL, J. P. The resolution of role conflict within the family. *Psychiatry*, 1957, **20**, 1-16.
- STERN, G. G. Environments for learning. In

- N. Sanford (Ed.), *The American college*. New York: Wiley, 1962. Pp. 690-730.
- STERN, G. G. Activities index and college characteristics index: Scoring instructions and college norms. Syracuse: Psychological Research Center, 1963.
- SUMMERSKILL, J. Dropouts from college. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1962. Pp. 627-657.
- SUPER, D. *Career development: Self-concept theory*. New York: College Entrance Examination Board, 1963.
- TUCKER, L. R. The extension of factor analysis to three-dimensional matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehart & Winston, 1964. Pp. 109-127.

(Received October 22, 1966)

DIFFERENTIAL MEMORY FOR PICTURE AND WORD STIMULI¹

JOSEPH R. JENKINS, DANIEL C. NEALE,

University of Minnesota

AND STANLEY L. DENO

University of Delaware

To compare the effects of using either pictures or words in a recognition task, 120 college sophomores were assigned randomly to 4 treatment conditions: (a) see pictures—recognize pictures (PP); (b) see words—recognize words (WW); (c) see pictures—recognize words (PW); (d) see words—recognize pictures (WP). When the number of correct identifications of the original stimuli were compared, Group PP was superior to WW ($p < .01$), and PW was virtually the same as WW but significantly superior to WP ($p < .025$). The tendency to admit instructions, identifying new stimuli as members of the initial series, was significantly more predominant in the WW condition than in the PP condition ($p < .001$). Results were discussed in relation to hypotheses about the nature of encoding processes for pictures and words.

The present study compared the effects of using either pictures or words in a recognition task.

A number of investigators have shown that pictures and words function differently as stimuli in learning tasks. Herman, Broussard, and Todd (1951), using either pictures or words to represent the same common objects, demonstrated that pictures are learned faster than words in a serial anticipation task. Lumsdaine (1949), Deno (1965), and Paivio and Yarmey (1966) all found that when pictures were used as stimuli in a paired-associate task, learning was faster than when words were the stimuli. Ducharme and Fraisse (1965), testing written recall of words and pictures, reported a tendency for children to recall more words than pictures but

for adults to recall more pictures than words. However, neither tendency was statistically reliable.

A difficulty exists, especially in serial anticipation and recall tasks, in contriving a fair measure of learning for the picture condition. In the above studies memory for pictures was measured by having the subject (*S*) say or write the *word* which labeled the picture. This procedure may have introduced an extra transformation for *Ss* who learned pictures compared to those who learned words.

The present study sought to provide a fairer measure of memory by using a recognition task in which stimuli were either words or line drawings representing common objects. Since in a recognition task the measure of memory is *S's* response, "Yes, I have seen it," or "No, I have not seen it," no advantage accrues to *Ss* whether the stimuli to be recognized are pictures or words.

In addition, the present study sought to test hypotheses about *Ss'* abilities to make the transition from

¹ This research was supported by grants to the University of Minnesota, Center for Research in Human Learning from the National Science Foundation (GS541), the National Institute of Child Health and Human Development (5-P01-HD-01136-03), and the Graduate School of the University of Minnesota.

picture to word stimuli and vice versa. Therefore, one condition was included in which Ss saw a set of pictures and then were asked to identify words which represented the pictures, and a second condition was included in which Ss saw a set of words and then were asked to identify the pictures. In this way the importance of changing stimulus modes could be assessed and information could be gained about how Ss encoded the two kinds of stimuli.

Several hypotheses guided the investigation. First, on the basis of previous research, pictures were expected to be more easily remembered than words. Therefore, the performance of a "see pictures—recognize pictures" group was expected to be superior to a "see words—recognize words" group.

Second, recognition was expected to be better when stimuli in the two presentations were in the same, rather than different, modes as in the initial presentation. However, for the highly labelable pictures used in the study, the tendency to encode pictures verbally was expected to be strong. Therefore, performance of a "see pictures—recognize words" group was expected to be equal, or nearly so, to that of a "see words—recognize words" group.

Third, a "see words—recognize pictures" group was expected to show poorest performance relative to other groups. Such a group would have the disadvantage of recognizing stimuli in a different mode without the help that verbal labeling gives to the "see pictures—recognize words" group. Presumably, the probability is remote that Ss seeing the words would encode them with an image to match the pictures in the experiment.

METHOD

The design of the study included four conditions: (a) See pictures—recognize

pictures (PP); (b) See words—recognize words (WW); (c) See pictures—recognize labels of the pictures (PW); (d) See words—recognize pictorial representations of the words (WP).

The Ss were 120 University of Minnesota male and female students enrolled in introductory psychology classes. The Ss were randomly assigned to one of the treatment conditions which were administered to groups varying in size from eight to 20. Each condition ultimately contained 30 Ss.

Stimuli were either pictures or words of 42 common objects. The pictures were simple black-white line drawings, chosen because they were easily and consistently labeled. Both pictures and words were separately reproduced on 35-mm slides. The slides were projected onto a movie screen by a Carousel projector adjusted with two Hunter timers so that stimulus presentation was 1.5 seconds per item with a 1.5-second interstimulus interval. Mimeographed data sheets were prepared for use by Ss in rating.

The experimenter (*E*) instructed Ss that they would see a series of pictures and words, according to the condition, and that they should remember them. In addition, Ss were cautioned that the order of presentation was not significant. Then *E* projected the 17 stimuli in the learning list (17 L) onto the screen. The 17 L included: BED, STAR, NAIL, RADIO, TABLE, TIE, NOSE, DOOR, PIE, CHURCH, CAT, WOMAN, BREAD, STAIRS, FISH, KNIFE, and BULB. Following their presentation, *E* distributed the mimeographed sheets and instructed the Ss that they would see a longer series of stimuli that would include those originally seen. The Ss were told to respond to each stimulus of the second series by marking a five-point rating scale accordingly: 1—the stimulus was definitely not from the original series; 2—I believe the stimulus was not from the original series; 3—I do not remember if the stimulus was in the original series; 4—I believe the stimulus was from the original series; 5—the stimulus was definitely in the original series. After an example the Ss saw 42 slides consisting of the 17 L slides and 25 additional slides (25 X) that were not included in the initial series. Those Ss in conditions that changed stimulus mode (PW and WP) were told that the second series would contain the original 17 L slides but that their form would be changed. This was followed by an example. Each group received the same random order of the 17 L slides, and also the same random order of the 42 test slides. The 17 L slides were associatively unrelated

while the 25 X slides were associatively related both to each other and to 17 L slides. For example, STAR was an original stimulus while MOON and SUN were members of the test list along with STAR.

RESULTS

Three specific comparisons were of primary interest: (a) PP versus WW; (b) WW versus PW; and (c) PW versus WP. The significance of difference in each of the contrasts was tested with a *t*-test.²

When the number of correct identifications of the 17 L stimuli (number of 4's and 5's for 17 L) are com-

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR
NUMBER OF CORRECT IDENTIFICATIONS
OF THE 17 L STIMULI

Group	<i>M</i>	<i>SD</i>	<i>t</i>
PP	15.73	1.17	2.48**
WW	14.23	2.16	0.06
PW	14.27	2.43	2.32*
WP	12.87	3.15	

Note.—MSE = 5.47.

* $p < .025$.

** $p < .01$.

pared (Table 1), PP is superior to WW, $p < .01$, and the mean for group PW is virtually the same as WW but significantly superior to WP, $p < .025$.

An investigation of the number of correct identifications for the 25 X stimuli (number of 1's and 2's for the 25 X stimuli) again demonstrates the superiority of the PP condition. The Ss in the PP group were able to re-

² A two-way analysis of variance might have been used, except that such a procedure would fail to provide direct tests of the hypotheses under consideration. Instead *t*-ratios were computed using in each denominator a pooled estimate of the variability over the four conditions. This quantity is equivalent to mean square error (MSE) in the analysis of variance. This procedure is similar to one suggested by Winer (1962, p. 69).

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR
NUMBER OF CORRECT IDENTIFICATIONS
OF THE 25 X STIMULI

Group	<i>M</i>	<i>SD</i>	<i>t</i>
PP	22.97	2.46	3.59**
WW	19.27	4.76	
PW	18.60	4.34	
WP	17.17	4.05	

Note.—MSE = 15.98.

* $.10 > p > .05$.

** $p < .001$.

spond "not there" more accurately than Ss in the WW group, $p < .001$. Again, conditions WW and PW were nearly identical in their performance though PW did not perform significantly better than WP, $.05 < p < .10$, the means fell in the predicted direction (see Table 2).

The tendency to admit intrusions, identifying 25 X stimuli as members of the initial series (number of 4's and 5's for 25 X), is significantly more predominant in the WW condition than in the PP condition, $p < .005$ (see Table 3). In line with the above findings, condition PW does not differ from WW and again is in the predicted direction relative to PW, $.05 < p < .01$.

A final analysis that compares the overall certainty with which Ss in the four conditions respond is summarized in Table 4. Group PP responded

TABLE 3

MEANS AND STANDARD DEVIATIONS FOR
NUMBER OF INCORRECT IDENTIFICATIONS OF
THE 25 X STIMULI (INTRUSIONS)

Group	<i>M</i>	<i>SD</i>	<i>t</i>
PP	1.03	1.59	3.01**
WW	2.93	2.45	
PW	3.50	2.89	1.59*
WP	4.50	2.64	

Note.—MSE = 5.96.

* $.10 > p > .05$.

** $p < .005$.

TABLE 4
MEANS AND STANDARD DEVIATIONS FOR
DEGREE OF CERTAINTY OF
IDENTIFICATIONS

Group	<i>M</i>	<i>SD</i>	<i>t</i>
PP	39.03	4.03	2.89*
WW	35.13	5.37	0.45
PW	34.53	5.59	0.89
WP	33.33	5.73	

Note.—MSE = 27.29.

* $p < .005$.

with absolute certainty (number of 1's and 5's for the entire 42 stimuli) significantly more frequently than group WW, $p < .005$.

DISCUSSION

The data confirm the hypothesis that pictures are more easily remembered than words. Several factors may contribute to this effect. Two possible explanations follow.

When a stimulus is encoded, both a nonlinguistic and a linguistic representation of the stimulus may be stored. The nonlinguistic representation is in some sense richer for pictures than for words. When the stimulus is a picture, for example, of BOY, distinctive cues may be present, perhaps a distinctive facial expression or distinctive clothing. These cues are irrelevant as far as identifying the stimulus as BOY, but nevertheless they enrich the nonlinguistic representation that is stored. This richness helps *S* to discriminate among picture stimuli. On the other hand, the word BOY has few distinctive characteristics that help *Ss* to discriminate among words. There is little reward for attending to the formal attributes of a word, for example, its shape, since these attributes infrequently aid discrimination in memory tasks such as those under consideration. In fact, the formal similarity of the words may be a source of interference (Underwood,

1953). In short, pictures contain distinctive cues which make them more discriminable than their labels and this discriminability enhances memory for pictures compared to their labels.

Even if one is reluctant to admit that a nonlinguistic representation is stored, it is likely that the additional cues are stored at least verbally, so that the picture BOY is encoded and stored as BOY—WITH BOOKS IN HAND, SMILING, LOOKS LIKE A SCHOOL BOY, WALKING, AGE APPROXIMATELY EIGHT. On the other hand, the additional cues for the word BOY are less conspicuous: BOY—PICA TYPE, ALL UPPER CASE LETTERS, THREE LETTERS. Moreover, the fact that the other stimuli have these same attributes only increases the similarity dimension when *Ss* are required to choose the original stimulus from the test list.

An alternative hypothesis relies on the associative characteristics of pictures and words. The same stimuli that were used in the present experiment were formerly employed in a comparison of free associations to pictures and words (Deno, Johnson, & Jenkins, 1966). The free-association data suggest that associations to a word are different from associations to its pictorial representation. Moreover, the general findings indicated that associative clumping, as measured by associative overlap, was of lesser magnitude between sets of words than between sets of pictures. Consequently, if presentation of a stimulus with instructions to remember elicits associations to that stimulus, then there is smaller chance that associative interference will accrue in the PP condition than in the WW condition. Unfortunately, an investigation of this hypothesis is extremely difficult with the present

data since Ss on the average missed less than two stimuli in condition PP.

Comparisons of the performance of groups PW and WW support the hypothesis that pictures are encoded and stored with their verbal label. In each comparison of the two groups, the means of the two groups are highly similar. Somewhat surprising is the apparent lack of a decrement in the performance of the PW group attributable to the change in stimulus mode. It may be concluded that recognizing the word labels of pictures is as easy as recognizing the words themselves. The Deno, Johnson, and Jenkins data suggest that the lower associative overlap between pictures as compared with words may also contribute to the performance of the PW group by a reduction in the associative interference during the learning stage. From this one would anticipate a reduction of the intrusions for the PW group in relation to the WW group, Table 3. The data do not confirm this expectancy. Apparently, if the associative factor contributes to the performance of any groups, both series must be in the same stimulus mode.

The WP condition had the greatest difficulty identifying the original stimuli. While the PW group apparently stored the stimulus verbally, the WP group had little chance to store the stimulus in the form in which it would be tested.

The results have some implications for foreign language instruction. Frequently, attempts are made to eliminate initial language habits by having students learn the meanings of foreign words through pairing these

words with pictures rather than their initial language equivalents. The assumption is that the learner will acquire the meaning of the foreign word directly as meaning was acquired within the native language. The data from the PW condition suggest, however, that Ss tend to encode a picture verbally within the native language anyway, so that attempts to eliminate previously established verbal habits may well be unsuccessful.

REFERENCES

- DENO, S. L. Multiple discrimination learning as a function of conceptual similarity among picture and word stimuli. Unpublished doctoral dissertation, University of Minnesota, 1965.
- DENO, S. L., JOHNSON, P. E., & JENKINS, J. R. The structure of associations to pictures and words. Paper presented at the meeting of the American Educational Research Association, Chicago, February 1966.
- DUCHARME, R., & FRAISSE, P. Etude génétique de la memorisation de mots et d'images. *Canadian Journal of Psychology*, 1965, **19**, 253-261.
- HERMAN, T., BROUSSARD, I. G., & TODD, H. R. Intertrial interval and the rate of learning serial order picture stimuli. *Journal of Genetic Psychology*, 1951, **45**, 245-254.
- LUMSDAINE, A. A. Effectiveness of pictures versus printed words in learning simple verbal associations. Unpublished doctoral dissertation, Stanford University, 1949.
- PAIVIO, A., & YARMEY, A. D. Pictures versus words as stimuli and responses in paired-associate learning. *Psychonomic Science*, 1966, **5**, 235-236.
- UNDERWOOD, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *Journal of Experimental Psychology*, 1953, **45**, 133-142.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received October 24, 1966)

EFFECTS OF CONCEPTUAL SIMILARITY ON SERIAL LEARNING AND RETENTION BY RETARDATES¹

ALFRED A. BAUMEISTER AND JUDITH GUFFIN

University of Alabama

2 groups of retardates learned 8-item serial lists under 2 conditions of item similarity. 1 list was composed entirely of pictures of animals. The other contained pictures of common objects from different conceptual categories. Association values were empirically matched for items in the 2 lists. Ss were run to a criterion of 1 perfect recitation of the items. After 48 hr. Ss were requested to relearn the same list. The results indicated that the conceptually related items were more difficult to learn. Relearning was not significantly affected by the treatments. Typical serial-position curves were found. The distribution of percentages of serial-position errors did not differ significantly over the 2 conditions.

Although intralist similarity is generally regarded as an important determinant of verbal learning rate (Underwood, 1963), the evidence for the effects of this variable in the performance of retardates is somewhat equivocal. Iscoe and Semler (Iscoe & Semler, 1964; Semler & Iscoe, 1965) have reported that, for retarded subjects (Ss), conceptually similar items in the paired-associate (PA) task are more difficult to learn and retain than dissimilar items. On the other hand, Wallace and Underwood (1964) found only a slight and nonsignificant difference in performance on PA lists of high and low similarity for retarded Ss. They speculated that less intelligent Ss are deficient in their ability to make appropriate implicit associative responses.

The major purpose of the present study was to determine whether conceptual similarity of items in a *serial* list would affect the learning and long-term retention of these items by institutionalized retardates. Serial-

learning rate should be particularly sensitive to this contextual variable. There is, moreover, some evidence to suggest that serial learning is relatively more difficult than PA learning for less intelligent individuals (Baumeister, 1967). An ancillary purpose was to examine serial-position effects in relation to intralist similarity.

METHOD

Subjects

Forty-eight Ss with IQs between 50 and 85 were obtained from a residential institution for the mentally retarded. All Ss with gross sensory or motor defects were excluded from the study. The mean IQ (Stanford-Binet) was 61.4; the mean chronological age was 21.7 years. The corresponding standard deviations were 8.1 and 4.7 years.

Learning Material

The test materials were selected from a set of 46 colored pictures of common objects. Preliminary to the experiment, association values were computed for each picture.² These values consisted of the median latencies to identify the object in the picture by 40 Ss from the population sampled in the main experiment. Two lists, each containing eight items, were composed from the parent set of 46 pictures. Items in the

¹Supported, in part, by United States Public Health Service grant HD-02588. The authors wish to express gratitude to the patients and staff at Partlow State Hospital and School for their cooperation.

²Gratitude is expressed to Norman Ellis for supplying the data from which these values were obtained.

two lists were matched closely on association values. One of the lists (high similarity) consisted exclusively of pictures of animals: rabbit, duck, fox, cow, sheep, tiger, monkey, and donkey. The other list (low similarity) comprised pictures of objects from different conceptual categories: train, flag, church, glove, fruit, grasshopper, pie, and ring.

Procedure

The Ss were randomly assigned to one of two subgroups. Tests of significance indicated no mental or chronological age bias in the composition of these subgroups. One group of Ss was required to learn the list composed of conceptually similar items and the other received the dissimilar items.

In order to minimize the effects of confounding of item position, the items in each list were systematically presented in orders following a Latin-square arrangement. Eight different orders, with three Ss under each, were employed, thereby insuring that every item appeared in every position an equal number of times.

The stimuli were presented on a screen using a Kodak Carousel projector. On the first presentation of the list, S was required to name the objects. Items were presented for 4 seconds. A blue slide appeared on the screen for 4 seconds between items. The S was instructed to anticipate the next picture during this interitem interval. An intertrial interval of 20 seconds was employed. The criterion of learning consisted of one perfect recitation of the items. Forty-eight hours later, S relearned the list to the same criterion.

RESULTS

The entire analysis was based upon errors. Summary data appear in Table 1. A *t*-test indicated ($p < .01$) that

the list composed of conceptually related items was more difficult to learn. No differences were found with respect to percentage savings in relearning the list after a 48-hour interval. A more detailed examination of the types of errors made under the two conditions indicated the list of similar items produced somewhat more frequent overt within-concept errors. However, proportions of intrusion and omission errors were not significantly different for the two groups. A statistical comparison of the serial-position curves under the two treatment conditions revealed no differences.

DISCUSSION

The results of the present experiment partially corroborate those reported earlier by Iscoe and Semler (Iscoe & Semler 1964; Semler & Iscoe, 1965) with respect to the detrimental effects of high conceptual similarity of items in structured verbal learning tasks. This effect seems to apply to the serial as well as the paired-associate learning of retarded Ss.

The theoretical analysis proposed by Wallace and Underwood (1964) would predict some of these findings. In brief, they have speculated that two implicit responses occur when a stimulus is presented in a verbal learning context. One of the responses involves the act of perceiving the

TABLE 1

MEAN ERRORS FOR SIMILAR AND DISSIMILAR CONDITIONS BY ITEM POSITION AND TOTALS FOR ORIGINAL LEARNING (OL) AND PERCENTAGE OF SAVINGS SCORES FOR RELEARNING (RL) AFTER 48 HOURS

Group	OL position								Total	RL percentage of savings
	1	2	3	4	5	6	7	8		
Similar	3.0	4.1	4.7	4.8	5.0	5.2	4.4	3.2	34.4	86
Dissimilar	2.3	2.5	2.9	3.1	3.6	3.8	3.3	1.9	23.4	80

Note.— $N = 24$ for each group.

stimulus and is described as a "representational response" (RR). The other covert response is said to be produced by the stimulus properties of the RR. This second response is labeled the "implicit associative response" (IAR) and consists of all the verbal associations called forth by the stimulus. Mediators are viewed as one type of IAR. When items which elicit the same IARs must be learned in some order, fewer differentiating cues are available and performance should be impaired. Thus a list composed of items from the same conceptual category should be more difficult to learn than a list of items from several different categories. The data obtained here are generally in agreement with this view, although one might have expected a greater proportion of within-concept intrusion errors under the condition of high similarity.

Relearning of the serial lists was not markedly different under conditions of high and low categorical similarity. However, little variability was noted in the relearning scores. Possibly, the effects of conceptual similarity upon retention are related to the degree of original learning. It should be noted that Semler and Iscoe (1965) found more long-term forgetting of paired associates when the PA sets were conceptually similar.

Although some wide interindividual differences were noted, typical group serial-position curves (nega-

tively skewed and bow-shaped) were found under both treatment conditions. The invariance principle regarding serial-position effects in this type task (McCrary & Hunter, 1953) appears to hold for contextual factors. Although increasing similarity of items affected overall difficulty, the distribution of percentages of serial-position errors did not differ significantly with respect to either bowing or skewing for the two treatment conditions.

REFERENCES

- BAUMEISTER, A. A. Learning abilities of the mentally retarded. In A. A. Baumeister (Ed.), *Mental retardation: Appraisal, education, and rehabilitation*. Chicago: Aldine Publishing Co., 1967. Pp. 181-211.
- ISCOE, I., & SEMLER, I. J. Paired-associate learning in normal and mentally retarded children as a function of four experimental conditions. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 387-392.
- MCCRARY, J. W., JR., & HUNTER, W. S. Serial position curves in verbal learning. *Science*, 1953, **117**, 131-134.
- SEMLER, I. J., & ISCOE, I. Concept interference and paired associates in retarded children. *Journal of Comparative and Physiological Psychology*, 1965, **60**, 465-466.
- UNDERWOOD, B. J. The representativeness of rote verbal learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964. Pp. 48-78.
- WALLACE, W. P., & UNDERWOOD, B. J. Implicit responses and the role of intralist similarity in verbal learning by normal and retarded subjects. *Journal of Educational Psychology*, 1964, **55**, 362-370.

(Received November 14, 1966)

INDIVIDUAL DIFFERENCES IN MEMORY SPAN WITH AND WITHOUT ACTIVITY INTERVENING BETWEEN PRESENTATION AND RECALL¹

ARTHUR E. WHIMBEY AND SANDRA L. LEIBLUM

University of Illinois

30 Ss were tested on 3 memory tasks. 1 of the tasks was a form of the traditional memory-span test. The other 2 tasks required the Ss to engage in some activity between the presentation and recall of the digit series. All 3 tasks were found to be reliable measures of individual differences. Furthermore, although the presence and type of activity intervening between presentation and recall was found to have a significant effect on the difficulty of the task, the intercorrelations of the 3 tasks were very high. It was concluded that interspersing activity between presentation and recall does not change the psychological processes involved, but only affects the difficulty of the tasks.

There has recently been renewed interest in short-term memory (STM) as an important human abilities variable. Ellis (1963) postulates that mental retardates are deficient in memory-trace strength. Baumeister and his co-workers (Baumeister & Bartlett, 1962, 1963) have found memory trace to be an important factor in the performance of retardates on the Wechsler Intelligence Scale for Children and have related memory span to learning tasks. Jensen (1964) pointed out that one of the main reasons that memory-span tests tend to have low correlations with other measures of intelligence is that most span tests are unreliable. He showed that the span tests of the Wechsler Adult Intelligence Scale (WAIS) correlated very highly with the WAIS vocabulary test when the correlation was corrected for attenuation.

Researchers in the area of verbal learning have studied STM extensively (e.g., Bruning & Schappe, 1965; Conrad, 1960; Peterson & Peterson, 1959) but have been only interested

in parameters that affect group means. Typically these studies require the subjects (Ss) to engage in some activity between presentations and recall in an attempt to prevent rehearsal. Most pertinent to the present study, Bruning and Schappe (1965) have shown that the amount of forgetting depends upon the type of intervening activity, and Conrad (1960) found that merely having *S* say the word "O" before trying to repeat a series of eight single-digit numbers resulted in a large decrement of percentage correct. However, little research has been performed investigating the stability of individual performance under different conditions of intervening activity. Individual differences in resistance to interference might actually be unrelated to individual differences in memory ability when no intervening stimuli are present; the intercorrelations among memory tasks involving different conditions of intervening activity might be quite low.

In view of the increasing importance attributed to STM by researchers interested in individual differences and the extensive amount of research being done on STM by learning psy-

¹This research was supported by a grant from the University Research Board, University of Illinois.

chologists, it seemed desirable that: (a) reliable tests of memory span be developed; (b) group-administered tests be devised; (c) an investigation be made of the relationship between individual differences in memory span with and without activity intervening between presentation and recall.

METHOD

Subjects

The Ss were 30 undergraduates who were paid for participating in the experiment.

Tasks

Each S was tested on three tasks. The tasks involved recalling series of single-digit numbers. The series were between 5 and 9 digits in length and were presented by a tape recorder at the rate of 2 digits per second. This rapid presentation rate prevented rehearsal or the use of other memory techniques between the presentation of each digit. In a previous unpublished study employing a 1 digit per 1.5 seconds rate, different Ss used different techniques, so that the scores measured individual differences in techniques as well as individual differences in memory span.

Each task consisted of 30 trials. In each block of five trials, one 5, 6, 7, 8, and 9 digit series were presented in random order. The three tasks differed in the activity which intervened between presentation and recall of the series.

Task 1. The Ss were instructed to attempt to write the series of digits as soon as the presentation was completed.

Task 2. In this task, Ss were told that after the series was presented, they were to first write the digit "0" and then try to write the digit series.

Task 3. For this task, Ss were given a sheet with the following word-letter (W-L) pairs: bear-Q, bird-L, crab-V, fish-H, frog-R, flea-J. The presentation of each series of digits was followed by one of the six words. The S was to look at the W-L sheet, write the letter corresponding to the word, and then attempt to write the digit series.

Procedure

The Ss were tested in groups of about 10. Three groups of Ss were used, and the order of presenting the three tasks was

counterbalanced across groups. A 10-minute rest break and a 10-minute syllogism test intervened between each task.

In scoring the tasks, a series was considered correct if it was reproduced in the order presented. The S's score was determined by allotting 5 points for each 5-digit series correct, 6 points for each 6-digit series correct, etc.

RESULTS

To evaluate the effect of the different intervening activities on the amount of recall and to assess the effects of practice, an analysis of variance was performed. Counterbalancing the presentation order of the three tasks for the three groups resulted in a Latin-square design. An equal number of Ss did not arrive at each of the three testing sessions. Furthermore, two Ss did not perform all three tasks properly (e.g., did not write "0"). To equalize the group frequencies, several Ss were randomly discarded, and the analysis was based on seven Ss per group.

The mean performance on the respective tasks were: Task 1, 137.2; Task 2, 107.6; Task 3, 58.3. The task effect was found to be significant ($F = 82.01$, $df = 2/36$, $p < .01$), but neither the practice effect ($F < 1$), the sequence effect ($F < 1$), nor the residual ($F < 1$) produced significant differences. Previous unpublished research based on the introspective reports of Ss suggested that writing "0" could be done automatically without interfering with subvocal rehearsal of the digit series. To determine whether writing "0" did have an effect, a test was made on the difference between the means on Task 1 and on Task 2. The difference was found to be significant ($F = 22.61$, $df = 1/36$, $p < .01$).

To determine whether the three tasks were measuring the same STM process, the intercorrelations of the tasks were obtained and are presented

TABLE 1
INTERCORRELATIONS AND RELIABILITIES OF
THE TASKS

Task	1	2	3
1	.913	.835	.803
2		.881	.735
3			.877

in Table 1. The r between Tasks 1 and 3 was based on 30 Ss; the other two r s were based on 29 Ss. An estimate of the reliability of each of the tasks was made by correlating the scores from Blocks 1, 3, and 5 with the scores from Blocks 2, 4, and 6, and then using the Spearman-Brown formula for doubled length. The resulting values are presented on the main diagonal of Table 1.

DISCUSSION

The significant effect of tasks found in the present experiment replicates previous findings (e.g., Conrad, 1960) that intervening activity tends to decrease recall. The technique of including a brief period of activity between the presentation and recall of some information has been used extensively since it was first introduced by Peterson and Peterson (1959). The important finding of the present experiment is that the three tasks correlated highly with each other (the correlations would be even higher if corrected for attenuation), despite the different intervening activities. Since Ss tended to maintain their relative positions on the tasks, the common practice of only comparing the mean performance on tasks involving different intervening activity is legitimate. It appears that interspersing activity between presentation and recall affects only the difficulty of the task but does not change the psychological processes involved.

If this were not true, a much more extensive analysis of the behavior would be necessary. Furthermore, a learning theory describing performance on these tasks would only need a level parameter to account for individual differences. Complicated task-by-subject functions would not be necessary.

Further support for the conclusion that the intervening activity affected only difficulty is provided by comparing the difference in difficulty level of pairs of tasks with the respective correlations. There is no trend for the correlation between tasks to decrease as the difference in difficulty increases (Tasks 1 and 3 are further apart in difficulty than Tasks 2 and 3, yet the r between Tasks 1 and 3 is .803, whereas the r between Tasks 2 and 3 is only .735).

The absence of a significant practice effect indicates that very little, if any, improvement with practice occurred in the experiment. Improvement with practice often indicates that S devised a strategy or technique to help him remember the material (Gates & Taylor, 1925). The absence of a significant effect suggests that it was impossible to develop elaborate recall strategies in the experimental tasks. The tasks assess short-term memory, not recall strategies.

Finally, it has been demonstrated that the measurement of memory span is not intrinsically unreliable. Good measures of span can even be obtained in a group-testing situation. The relationship between memory span and other general intelligence variables should now be obtained directly, rather than being estimated by correction formulas. To whatever degree memory span is found to relate to other psychometric measures, the constructs and relationships

discovered in experimental learning studies should effect performance on these measures.

REFERENCES

- BAUMEISTER, A. A., & BARTLETT, C. J. Further factorial investigations of WISC performances of mental defectives. *American Journal of Mental Deficiency*, 1962, **67**, 257-261.
- BAUMEISTER, A. A., BARTLETT, C. J., & HAWKINS, W. F. Stimulus trace as a predictor of performance. *American Journal of Mental Deficiency*, 1963, **67**, 726-729.
- BRUNING, J. L., & SCHAPPE, R. H. Type of interpolated activity and short-time memory. *Psychological Reports*, 1965, **16**, 925-929.
- CONRAD, R. Very brief delay of immediate recall. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 45-47.
- ELLIS, N. R. The stimulus trace and behavioral inadequacy. In N. R. Ellis (Ed.), *Handbook of mental deficiency*, New York: McGraw-Hill, 1963. Pp. 134-158.
- GATES, A. I., & TAYLOR, G. A. An experimental study of the nature of improvement resulting from practice in a mental function. *Journal of Educational Psychology*, 1925, **16**, 533-592.
- JENSEN, A. R. Individual differences in learning: Interference factor. *Cooperation Research Report #1867*, 1964. Department of Health, Education and Welfare.
- PETERSON, L. R., & PETERSON, M. J. Short-term of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193-198.

(Received November 14, 1966)

New and Important From HRW

THE PROCESS OF SCHOOLING

A Psychological Examination

J. M. STEPHENS, University of British Columbia

Professor Stephens explores the phenomenon of learning as a process that somehow continues in spite of innovations and changes. This point of view represents the author's belief that education is a natural process which teachers must try to abet as well as they can—a position which is intended to provoke some different, new ways of looking at the educational work teachers are trying to do. To this end, the author investigates the causes for the many disappointing results that so often follow administrative innovations.

July 1967

176 pp.

\$3.95 paper

By the same author

THE PSYCHOLOGY OF CLASSROOM LEARNING

Drawing upon a profusion of recent data, Professor Stephens integrates new findings with traditional knowledge through detailed application of both to specific classroom situations. "Informational boxes" summarize investigations into the nature and process of learning and into factors that influence learning. Copious suggestions for further reading, discussion questions, an appendix on statistics, and indexes further complement the text.

1965

525 pp.

\$8.95

Instructor's Manual

THE PSYCHOLOGY OF LANGUAGE, THOUGHT, AND INSTRUCTION Readings

Edited by JOHN P. De CECCO, San Francisco State College

Assuming the central importance of language to intelligent human thought and of language and thought to effective instruction and learning, Professor De Cecco has drawn together in an ambitious synthesis research articles from linguistics, psycholinguistics, cognitive and developmental psychology, language behavior, and articles that attempt at least tentative applications of these other concerns to the planning of instruction.

1967

464 pp.

\$7.95

Holt, Rinehart and Winston, Inc.

383 Madison Avenue, New York, N.Y. 10017



EDUCATIONAL PSYCHOLOGY

An Application of Social and Behavioral Theory

by LOUIS M. SMITH and BRYCE B. HUDGINS,
both of Washington University, St. Louis
Knopf, 576 pages; 21 figures; 73 tables; \$8.50

This is a comprehensive analysis of the principles and substantive findings of educational psychology. While maintaining strong emphasis on the practical problems which confront every prospective teacher, the authors present the most significant educational theories and research in recent times. An INSTRUCTORS MANUAL is available.

PSYCHOLOGY OF HUMAN ADJUSTMENT

by LESTER D. CROW, *Brooklyn College*
Knopf, 1967; 640 pages; \$7.95

Probes the nature and problems of personal and social adjustment. This text defines such adjustment problems as frustration and conflict, defense mechanisms, and peer-age and adult relationships. An INSTRUCTORS MANUAL is available.

In preparation. . .

MENTAL HEALTH IN THE SCHOOLS

by THOMAS A. RINGNESS, *University of Wisconsin*
Random House, December 1967; 512 pages; \$6.95

A group approach in which "mental health" problems are treated as essentially problems in learning to be coped with through learning principles.

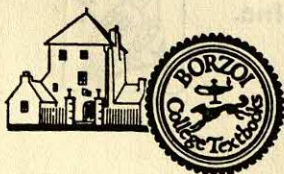
Now available. . .

SOCIETY'S CHILDREN

A Study in Ressentiment in Secondary School Education

by CARL NORDSTROM, *Brooklyn College*,
EDGAR Z. FRIEDENBERG, *SUNY, at Buffalo*
and HILARY A. GOLD, *Brooklyn College*
Random House, July 1967; SED 7; 224 pages; \$2.45 paperbound

An incisive view of the American secondary school today, which reveals how individuality and creativity are stifled, how enthusiasm is suppressed, and how conformity is perpetuated. Comprehensive analysis is based on findings in nine secondary schools.



RANDOM HOUSE, Inc.
ALFRED A. KNOPF, Inc.

The College Department
501 Madison Avenue
New York 10022

The Psychological Foundations of Education Series

Edited by Victor H. Noll, Michigan State University

Intensive investigation in the field of educational psychology has resulted in the recognition of a wide variety of significant topics that warrant independent consideration. The *Psychological Foundations of Education Series* is designed to provide a deeper examination of specific areas which are not given thorough treatment in the many comprehensive textbooks now in use. The volumes, combining theoretical background material with practical classroom applications, will increase the impact and effectiveness of current courses in the psychological foundations sequence.

The Psychology of Adolescence for Teachers

By Glenn Myers Blair, and R. Stewart Jones, both of University of Illinois
1964, 128 pages, paper, \$1.50

The Psychology of the Child in the Classroom

By Don C. Charles, Iowa State University
1964, 96 pages, paper, \$1.25

The Psychology of Learning in the Classroom

By Robert C. Craig, Michigan State University
1966, 85 pages, paper, \$1.25

The Mentally Retarded Child in the Classroom

By Marion J. Erickson, Ypsilanti Public Schools
1965, 128 pages, paper, \$1.25

The Psychology of Discipline in the Classroom

By William J. Gnagey, Illinois State University
1968, approx. 80 pages, paper, prob. \$1.50

Problem Solving in the Classroom

By Bryce B. Hudgins, Washington University
1966, 96 pages, paper, \$1.25

Teacher Self-Evaluation

By Ray H. Simpson, University of Illinois
1966, 96 pages, paper, \$1.25

Guidance in the Classroom

By Ruth Strang, University of Arizona, and Glyn Morris, Assistant Supervisor of Guidance and Curriculum, Lewis County, New York
1964, 128 pages, paper \$1.50

Gifted Children in the Classroom

By E. Paul Torrance, University of Georgia
1965, 112 pages, paper, \$1.25

Write to the Faculty Service Desk for examination copies.

THE MACMILLAN COMPANY

866 Third Avenue, New York, New York 10022

ALLYN AND BACON, INC.

EVALUATING PUPIL GROWTH: Principles of Tests and Measurement, III

J. Stanley Ahmann, Colorado State University and **Marvin D. Glock**, Cornell University. A balanced treatment of teacher-made tests and standardized instruments. The role of educational objectives has been strengthened. Consideration is given to taxonomies of educational objectives stated in specific behavioral terms. New 1967. 605 pp.

THE CAUSES OF BEHAVIOR: Readings in Child Development and Educational Psychology, II

Judy F. Rosenblith, Wheaton College and **Wesley Allin-smith**, University of Cincinnati. Designed for the increasingly frequent courses of study that combine human development and educational psychology, this broad collection of recent articles covers the scope of present knowledge and the range of theory and method in psychology today. The second edition has been carefully brought up to date to include the latest research studies in child development and educational psychology.

1966

Paperbound

608 pp.

EDUCATIONAL PSYCHOLOGY, II

James M. Sawrey and **Charles Telford**, both of San Jose State College.

STUDENT GUIDE TO ACCOMPANY EDUCATIONAL PSYCHOLOGY, II

Earl L. Jandron, San Jose State College.

AN INTRODUCTION TO STATISTICS AND MEASUREMENT: A Programmed Book

Lowell A. Schoer, State University of Iowa.

HUMAN DEVELOPMENT IN WESTERN CULTURE, II

Harold W. Bernard, Oregon State System of Higher Education.

READINGS IN HUMAN DEVELOPMENT

Harold W. Bernard and **Wesley C. Huckins**, both of the Oregon State System of Higher Education.

470 Atlantic Avenue • Boston, Mass. 02210

Announcing

The New **Otis-Lennon Mental Ability Test**

*a completely new edition of the well-known
Otis Quick Scoring Mental Ability Test series*

The various levels of the **Otis-Lennon Mental Ability Test** series are designed to provide comprehensive and continuous assessment of the general mental ability or scholastic aptitude of pupils in grades K-12.

In addition to retaining many of the outstanding features of previous editions, this entirely new series incorporates further refinements in item content, organization of the various levels, and scaling and norming procedures. These improvements were specifically designed to guarantee a continuous, carefully articulated scale for the measurement of general mental ability.

Now, for the first time, a combined scoring and reporting service is being offered for two of our most popular test offerings—**Stanford Achievement Test** and **Otis-Lennon Mental Ability Test**. This service, available for grades 4-12, is designed to provide the teacher and administrator with a single report containing the results for the two test services.

Use of a single MRC Answer Folder facilitates administration of the tests and reduces the need for later collating of test-score data resulting from the use of a separate answer document for each test.

*For further information on Otis-Lennon Mental Ability Test and on the combined Achievement/Mental Ability Service Program write:
Programs and Services Division*



Harcourt, Brace & World, Inc.

TEST DEPARTMENT

757 THIRD AVENUE, NEW YORK, N. Y. 10017

Chicago • San Francisco • Atlanta • Dallas

PROCEEDINGS

of the 75th Annual Convention of the American Psychological Association 1967

The purpose of the **PROCEEDINGS** is to facilitate the dissemination of information regarding current work in major fields of psychology by making available research reports with a minimum of lag and well in advance of the Convention. The 1967 **PROCEEDINGS**, published in July, contains approximately 200 reports distributed across the major subject areas of psychology.

The 1967 Convention **PROCEEDINGS** will contain the contributed papers of the following Divisions:

- | | |
|--|---|
| 1. <i>Division of General Psychology</i> | 14. <i>Division of Industrial Psychology</i> |
| 3. <i>Division of Experimental Psychology</i> | 15. <i>Division of Educational Psychology</i> |
| 6. <i>Division of Physiological and Comparative Psychology</i> | 16. <i>Division of School Psychologists</i> |
| 7. <i>Division of Developmental Psychology</i> | 17. <i>Division of Counseling Psychology</i> |
| 12. <i>Division of Clinical Psychology</i> | 19. <i>Division of Military Psychology</i> |

PRICE: \$9.00 per copy (\$4.50 to APA members)

Copies of the 1966 **PROCEEDINGS** are still available at \$6.00 per copy (\$3.00 to APA members)

AMERICAN PSYCHOLOGICAL ASSOCIATION

1200 17th Street, N.W.

Washington, D.C. 20036

PLEASE SEND ME _____ copy(ies) of the 1967 PROCEEDINGS

☐ Remittance Enclosed ☐ Bill Me (when issue mailed)

NAME _____

ADDRESS _____

Zip Code _____

The 1966 **PROCEEDINGS** was the first issue as a permanent publication. The 1965 issue was part of a research project on Scientific Information Exchange in Psychology.

JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 58

December 1967

Number 6, Part 1 of 2 Parts

CONTENTS

<i>Pitch Discrimination among Primary School Children</i>	315
ORPHA K. DUELL AND RICHARD C. ANDERSON	
<i>Transfer of Solution Rules in Problem Solving</i>	319
FRANCIS J. DI VESTA AND RICHARD T. WALLS	
<i>Two Methods for Adapting Self-Instructional Materials to Individual Differences</i>	327
RALPH J. MELARAGNO	
<i>Teacher Verbal Cues and Pupil Performance on a Group Reading Test</i>	332
GEORGE S. LAMB	
<i>Attentional Process in Reading: The Effect of Pictures on the Acquisition of Reading Responses</i>	337
S. JAY SAMUELS	
<i>Prediction of Student Accomplishment in College</i>	343
JAMES M. RICHARDS, JR., JOHN L. HOLLAND, AND SANDRA W. LUTZ	
<i>Learning Verbal and Symbolic Statements of Mathematical Rules</i>	356
JOSEPH M. SCANDURA	
<i>Meaning and Transfer of Verbal Learning</i>	365
JULIUS M. SASSENATH AND GEORGE D. YONGE	
<i>Reading Rate and Immediate versus Delayed Retention</i>	373
STANTON P. THALBERG	
<i>Perception of Deviant Behavior: A Study of Cheating</i>	379
JAMES Q. KNOWLTON AND LEO A. HAMERLYNCK	
<i>Cooperative versus Competitive Discussion Methods in Teaching Introductory Psychology</i>	386
DONALD BRUCE HAINES AND W. J. MCKEACHIE	

This is the last issue of Volume 58.

Volume Title Page and Contents of the Volume appear herein.

PITCH DISCRIMINATION AMONG PRIMARY SCHOOL CHILDREN¹

ORPHA K. DUELL AND RICHARD C. ANDERSON

University of Illinois

The pitch discrimination performance of 168 1st, 2nd, and 3rd graders was investigated in a natural classroom setting. Ss judged pairs of pure tones presented by a tape recorder as either the same or different. The standard ranged from 390 to 440 cps while the interval between the standard and the comparison ranged from $\frac{1}{8}$ of a half-step to a major sixth. 68% of the Ss discriminated intervals as large as and larger than a half-step; however, 4% did not discriminate differences as large as a sixth. Performance improved from the 1st to the 3rd grade. Performance did not change within the 30-minute testing period.

The ability to discriminate between different pitches is recognized as necessary for the appreciation and production of music. Much attention has been devoted by psychophysicists to the measurement of a difference limen (DL), that smallest pitch difference a person can discriminate accurately and reliably, for various areas along the pitch continuum for adult subjects (Ss). No such exhaustive effort has been made for younger Ss. Books for the preparation of future music teachers in elementary schools admonish them to draw attention to the direction of melodic movement and at least one book (Bergethon & Boardman, 1963) suggests that by the end of the second grade students should be able to "identify melodic movement as moving by steps and skips." Using the scale as a reference point this would imply that Ss are expected to discriminate differences as small as a half-step, which would be well within their ability if one generalizes from adult data; however, data from a

study conducted in Russia (Repina, 1961b) suggest that Ss of this age might not discriminate intervals of this size. Repina (1961b) measured the DL for a group of 40 preschool Ss ages 3 to 7, of which 10 were in the 6- to 7-year-old group. Of these 10 Ss one was reported to have a threshold "greater than a sixth," four had a threshold of a minor third to a perfect fourth, and five had a threshold of a half-step to a major second. These DLs were obtained using piano tones whereas Repina (1961a) found even larger DLs when using pure tones, which are generally used in adult DL studies, with a different group of eight Ss.

The purpose of the present study was to discover how large a pitch interval primary school children differentiate in the natural classroom setting, since this is the setting in which they produce and listen to music. If the Repina (1961a, 1961b) data are representative of young children's pitch discrimination performance, then school music teachers should not assume that children can already hear pitch differences of the size found in music (the half-step) and should concentrate on improving their students' abilities to hear such differences before expecting

¹This report is based upon a master's thesis submitted by the first author and directed by the second. We are indebted to Gordon Greenberg for advice and for assistance recording the tapes and to Robert Stake for statistical advice.

them to enjoy listening to and singing songs containing such intervals.

METHOD

Sample

The Ss were 168 children from the first, second, and third grades in a midwestern school. None of the Ss had been identified in the regular school testing program as suffering from hearing loss.

Equipment

The stimuli were pure tones generated by oscillators, recorded on tape, and presented to the Ss by means of a Wollensak Model T1616 tape recorder augmented with a 12-inch external speaker. It is recognized that by the time the sounds were recorded and released in the room, they could no longer be considered "pure." No attempt was made to employ a standard intensity level since a measure taken at 7 feet from the speaker would yield one reading while a measure taken at 8 feet in the same direction would yield a different reading. The volume control was set so that the stimuli could be easily heard in all areas where Ss sat. The intensity level was probably somewhere in the range of 50-60 decibels sound pressure level. All groups of Ss were tested in the same room with a maximum of 31 Ss per group.

Procedure

The method for presenting the stimuli was one in which the comparison tones are randomly arranged and the standard gradually changed through a small range. This method was found to be the most sensitive of 17 different methods compared by Harris (1949). The first step in this procedure is to choose a range within which to vary the standard (the tone to which the other tone, the comparison, is to be compared) that is small enough so that one can assume that the DL will not be different at the two extremes. Harris (1949) varied his standard from 750 to 850 cps, a range of 100 cps. The range of interest in the present experiment was the singing range of the Ss, middle C (261 cps) to the E an octave and a third above middle C (659 cps). The middle range within this broader range was explored. A 100 cps range would not have been appropriate at this point in the pitch continuum since the DL measured in cps becomes progressively smaller as one moves down the pitch continuum. Since Harris's

range included three half-steps, a cps range of three half-steps from 390 cps to 440 cps was covered. The standard systematically began at the lowest extreme, increased five cps on each comparison until the upper limit was reached, then descended by five cps steps until the lower limit was again reached. This cycle continued until all comparisons had been made. The comparison tones, both those above and below the standard, were arranged in random order and each interval was calculated from the standard with which it was paired. All intervals in this study were calculated according to the formula for the even-tempered scale, which is the tuning most generally used.

Following Repina (1961a, 1961b), in addition to the unison, the intervals that were investigated were the major sixth, the perfect fourth, the minor third, a half-step, two-thirds of a half-step, and one-third of a half-step. Each of the intervals was presented four times above the standard and four times below the standard. The unison was presented 16 times so that "same" was the correct response one-third as many times as "different" was the correct response.

Four orders of the comparisons were made so that each order contained every comparison once. The comparisons were randomized within each order. After each order was presented Ss stood and stretched for about 2 minutes. It took approximately 30 minutes to give the directions and complete the test.

Each stimulus presentation began with .25 seconds of white noise which informed the experimenter (*E*) that a new presentation was beginning. Then, in order, came 4 seconds of silence, during which *E* stated the number of the presentation; .75 seconds of the standard tone; .25 seconds of silence; .75 seconds of the comparison tone; 5 seconds of silence during which Ss marked their answers; and then .25 seconds of white noise that ended one presentation and signaled the beginning of the next. The tape recorder was equipped with a foot pedal which enabled *E* to stop the tape between presentations if necessary to replace broken pencils and to insure that all Ss were on the same question.

In an earlier experiment (Duell, 1965), Ss judged the second tone as higher, lower, or the same as the first tone in a pair. First graders performed at chance level on this task, perhaps because they were confused about the concepts higher and lower. In this study Ss judged the pair as either the same or different. They were supplied with answer sheets, which contained for each presentation

the alternatives D for different and S for same. General instructions were given and sample tones were sounded. Two examples (major sixth above and unison) were then presented with the correct answer announced before each, followed by three examples (unison, major sixth above, and perfect fifth above) with the correct answer given after each example. The Ss were warned that some of the differences would be very small.

All Ss marked the examples correctly and all made some response to the majority of presentations. Observation suggested that all or almost all worked diligently at the task. There were few signs of inattention even at the end of the testing period.

RESULTS AND DISCUSSION

Figure 1 traces the pitch discrimination performance for each grade. The means for the unison for Grades 1, 2, and 3 were 85%, 83%, and 96%. Analyses of variance indicated that there were significant differences ($\alpha = .01$ for all tests mentioned herein) among intervals and among grades. As can be seen in Figure 1,

TABLE 1
CUMULATIVE PERCENTAGE OF Ss WITH THRESHOLDS OF A GIVEN SIZE

Interval	Grade			
	First	Second	Third	All grades
Sixth	94	94	100	96
Fourth	91	91	100	94
Sharpened second	78	86	95	86
Half-step	59	66	80	68
Two-thirds half-step	36	40	60	45
One-third half-step	6	12	20	13

Note.—These percentages are based upon mean DLs obtained by combining the UDLs and LDLs. *N* for the first grade = 49; for the second grade = 58; for the third grade = 61. Total *N* = 168.

the percentage of correct discriminations increased as the size of the interval increased and improved from Grade 1 to Grade 3. The *F*s for or-

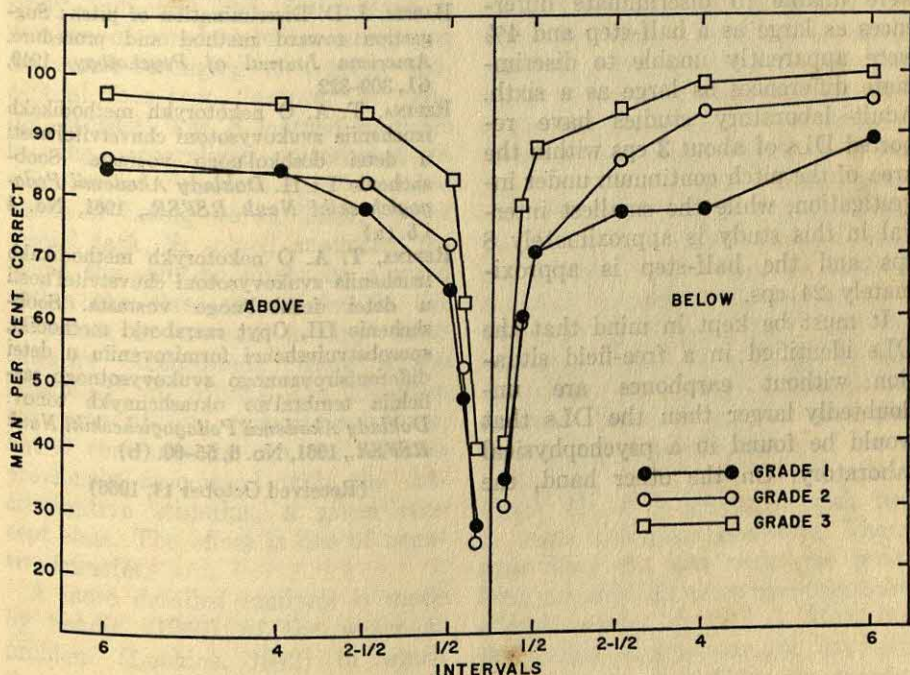


FIG. 1. Discrimination as a function of the size of the interval.

der were all less than 1.00, indicating neither a practice effect nor a loss of efficiency due to fatigue as testing progressed, or perhaps these factors canceled each other out. The F s for the Order $\times$ Interval interactions were significant; however, graphs revealed no trends for any of the groups. Intervals were slightly but not significantly easier to discriminate when the comparison tone was lower than the standard.

The principal goal of this study was to determine the size of the pitch interval which children will discriminate in a classroom milieu. The DL was defined as the smallest interval on which three of the four discriminations were correct and which was followed by an interval in which three out of four discriminations were also correct. Table 1 contains the cumulative percentage of children who obtained DLs of a given size. About 30% of the children tested were unable to discriminate differences as large as a half-step and 4% were apparently unable to discriminate differences as large as a sixth. Adult laboratory studies have reported DLs of about 3 cps within the area of the pitch continuum under investigation, while the smallest interval in this study is approximately 8 cps and the half-step is approximately 24 cps.

It must be kept in mind that the DLs identified in a free-field situation without earphones are undoubtedly larger than the DLs that would be found in a psychophysical laboratory. On the other hand, the

procedures used in the present study do resemble the conditions under which school music programs are conducted. Whether results entailing "pure" tones can be generalized to situations where complex tones are used is a moot question, and whether data gathered using any one type of complex tone would be representative of various complex tones is unanswerable at present. Nonetheless, until such questions are answered, the present study raises doubts about the value of primary school music programs for some children, and suggests a role for pitch discrimination training if effective techniques can be devised.

REFERENCES

- BERGETHON, B., & BROADMAN, E. *Musical growth in the elementary school*. New York: Holt, Rinehart and Winston, 1963.
- DUELL, O. K. Pitch discrimination among primary school children. Unpublished masters thesis, University of Illinois, 1965.
- HARRIS, J. D. Discrimination of pitch: Suggestion toward method and procedure. *American Journal of Psychology*, 1949, 61, 309-322.
- REPINA, T. A. O nekotorykh metodikakh izucheniiia zvukovysotoni chuvstvitel'nosti u detei doskol'nogo vozrasta. Soobshchenie I i II. *Doklady Akademii Pedagogicheskikh Nauk RSFSR*, 1961, No. 4 i 5. (a)
- REPINA, T. A. O nekotorykh metodikakh izucheniiia zvukovysotoni chuvstvitel'nosti u detei doskol'nogo vozrasta. Soobshchenie III. Opyt razrabotki metodiki, sposobstvuiushchei formirovaniu u detei differentsirovannogo zvukovysotnogo razlichiiia tembral'no okrashennykh tonov. *Doklady Akademii Pedagogicheskikh Nauk RSFSR*, 1961, No. 6, 55-60. (b)

(Received October 14, 1966)

TRANSFER OF SOLUTION RULES IN PROBLEM SOLVING

FRANCIS J. DI VESTA AND RICHARD T. WALLS

Pennsylvania State University

The effects of 5 stimulus-response relationships between Tasks A and B on the transfer of solution rules in problem solving were investigated. Facilitation was observed where solution rules and their applications were similar in the 2 tasks. Less facilitation occurred when the associations of solution rules with discriminative stimuli were interchanged in Task B. Least facilitation was found in the control groups who learned solution rules in Task A that were inappropriate for the solution of problems in Task B. 3 levels of complexity in the application of rules to the solution of problems were identified.

The present experiment was contrived to represent certain assumptions about problem solving posited by Schulz (1960). In particular we were interested in his analysis of such tasks as the water jar and anagram problems. These ostensibly unitary tasks may be analyzed into two functionally separable parts closely resembling the transfer paradigm.

Thus, in Rees and Israel's (1935) study, subjects (Ss) first solved a number of single-solution anagrams for words belonging to a given conceptual class (e.g., "nature," or "eating" words). The S learned to react to the generalized discriminative stimulus of "concept-class"—nature words or eating words. Then, in the second task, Ss solved anagrams for which the solution might be a word within the class represented in the first task or a word in any other category. Accordingly, there is the possibility of solving the anagram with one of two response-solutions; but S continues to use the solution previously associated with the discriminative stimulus, a given concept-class. The effect is one of negative transfer.

A more detailed analysis is made by Schulz (1960) of the water-jar problem (Luchins, 1942) in which the transfer processes parallel those

found in the anagram problems. In the first part of the task (Task I) S is presented with six problems. Each problem contains a stable stimulus pattern of three water jars and a stable response-solution (e.g., B-A-2C) or solution rule. The set of water jars can be considered a discriminative stimulus, that is, a cue for a specific solution rule, since three water jars are common to all problems. As a result of problem variation, but always with three water jars, the discriminative stimulus becomes generalized and associated with a given solution rule. With repeated presentations of problems, in which the three water jars are always present, the discriminative stimulus gains increased associative strength with the solution rule. Thus, in Task I, the most consistent S-R association is that between "water jars" and the solution rule of B-A-2C. It is this association that gains the most habit-strength, in the classic example, through six repeated trials.

In the second part of the task (Task II), S is presented with two or more additional problems. These, since they are also water-jar problems, embody the same discriminative stimuli present in Task I. However, the correct response solution has been changed from B-A-2C to the simpler

A+C or A-C solutions. Accordingly, *S* tends to employ the B-A-2C solution, the stronger habit resulting from association with the general water-jar stimuli (not the size of the jars) in the six previous problems. As in the above analysis of the anagram problem, the effect is one of negative transfer.

It is apparent from this analysis that the water-jar and anagram tasks are not only comprised of two parts, but they also involve nonspecific or generalized discriminative stimuli that are associated with response solutions. These relationships clearly parallel those found in transfer paradigms. Whether positive or negative transfer is obtained depends, in part at least, on the relationship of the nonspecific discriminative stimulus (*S*) and solution rule (*R*) in the initial task to the *S-R* relationships in the second task.

In this context, the notions of "rigidity," "set," and the like, commonly used to describe the effect obtained with the water-jar, anagram, functional fixedness, and "Umweg" problem-solving tasks may represent instances of negative transfer. An analysis of the task into its component parts suggests that positive transfer, in addition to negative transfer effects, may be observed in the situation. If these tasks are viewed as being comprised of a single series of discrete problems there is the possibility that one may be "blinded" not only to the existence of the original task, transfer-task components but also to the possible existence of an effective or functional nonspecific discriminative stimulus. Perhaps as a consequence of this neglect only one of a number of possible transfer paradigms was investigated in which the actual emphasis was on manipulations only of the solution rules or

strategies—although ostensibly both stimuli and responses were changed.

The analysis of transfer processes in verbal learning has demonstrated that many variations of such stimulus-response (*S-R*) relationships between the Task I and Task II phases of learning are particularly influential factors in determining the amount and direction of transfer. The A-B_r paradigm, for example, has been extensively investigated and shown to produce massive amounts of interference (Twedt & Underwood, 1959). In this paradigm, *Ss* learn several (A-B) pairs of words such as "house-young" and "certain-while" during Task I; then, in Task II, the combinations are re-paired to form (A-B_r) combinations such as "house-while" and "certain-young." Several additional paradigms can be employed to examine transfer produced by other variations of *S-R* combinations. These paradigms can be designated by the *S-R* terms in Task II learning (where the terms in Task I are always designated as A-B) as follows: The facilitation paradigms may be of the form A-B or A-B' where A and B are identical in both tasks and where B' may be an associate or otherwise similar to the B, or response term, used in Task I. Among the interference paradigms are those labeled C-D (both the stimulus and response terms in Task II are different from those in Task I); A-D (stimulus terms are the same in the two tasks but responses are different); C-B (stimulus terms differ in the two tasks and response terms are the same); and A-B_r (stimulus and response terms are the same in both tasks but are re-paired in Task II).

The foregoing paradigms have been described in detail by Underwood (1966), who also indicated that they

have been applied primarily in verbal learning because of the analytical flexibility of such tasks as paired-associate learning. However, the effect of these relationships on problem solving has not been extensively investigated (Di Vesta & Walls, 1967). As Schulz (1960) concludes,

It is...possible that if some of the complex problems of everyday life could be analyzed into their S-R components, it would be found that the relationships [just described] would [also] be characteristic of the most difficult problems. Thus, suppose we have trained our students to make a particular set of responses to a particular set of stimuli. Later, the student finds himself confronted with a problem which contains this set of stimuli. However, for a variety of reasons (e.g., errors in original training, errors in original learning, new discoveries, etc.) the S-R pairings required for solution of this problem are different from those required previously. How can we minimize the amount of negative transfer this student is likely to encounter [p. 75]?

This orientation, expanded further in the discussion section of this paper, is the basis of the present study.

In summary, the purpose of the present experiment was to examine the relative effects of the several S-R relationships, described at a molar behavioral level, between the Task I and Task II phases of anagram solution. It was assumed that initial (Task I) experiences in problem solving would facilitate or hinder performance in new (Task II) problem-solving situations in much the same way as these effects are observed in paired-associate learning. However, the emphasis in the present study is on generalized transfer, that is, on transfer associated with manipulations of discriminative stimuli and solution rules, rather than on the transfer of specific stimulus and response elements, respectively. Thus, it was hypothesized that when experiences with the application of solution

rules in Task I are congruent with or similar to the relationships required for effective solution of problems in Task II, the solution of anagrams in Task II will be facilitated. Conversely, where the relationships between the application of the solution rule in Task I and Task II are changed, the solution of new anagrams will be hindered in Task II.

METHOD

Subjects

The Ss were 75 undergraduate volunteers from the introductory educational psychology course at Pennsylvania State University. There were 27 males and 48 females. The Ss were assigned, by reference to a table of random digits, to one of the five experimental groups. Randomization was restricted to the set of experimental conditions and recycled at $N + 1$ treatment assignments. Although some Ss had participated in verbal learning experiments, only one had experience with anagram problems in experimental settings. The data for that S was eliminated from the analysis.

Design

The overall design was patterned after that used by Jenkins, Foss, and Odom (1965). Each S participated in two tasks. Task II was the transfer task and was common to all experimental conditions, hence, it shall be labeled A-B for convenience in presentation. In Task II S solved 16 eight-letter anagrams. Each anagram was preceded by one of four discriminative stimuli, that is, by an asterisk, number symbol, ampersand, or quotation mark. In Task I S practiced solving eight-letter anagrams for nonsense words by associating one of several solutions with a given discriminative stimulus until each solution could be used without direct reference. As in Task II, a discriminative stimulus always preceded the eight-letter anagram.

The manipulations of the relationship between the discriminative stimulus and response solution in the two tasks were accomplished in Task I. The following paradigms, labeled to parallel those employed by Jenkins et al. (1965), comprised the experimental conditions; associative facilitation (A-B'), warm-up (D-E), classical inter-

ference (A-E), associative interference (A-B'), and re-paired control (A-B). The overall design was thus represented by five interference-facilitation levels in which the assumed relationship between Task I and Task II is one of judged rather than of observed similarity or dissimilarity.

Conditions

Associative facilitation. In Task I (A-B') this group learned four S (discriminative stimulus)-R (solution rule) pairs. The solution rule was always in verbal form and the symbols were different from the number symbols used in Task II. Thus, the solutions in the two tasks were related but not identical. Each pair was learned in conjunction with the actual solution of four different anagrams for nonsense words. An example was the stimulus

- + - + - + - +
* E I N O Z P U H

for which the printed rule was the verbal statement, "All the minus letters and then all the plus letters." The S's task was to learn to spell the nonsense word, "E-N-Z-U-I-O-P-H," before looking at the rule printed on a separate card. In Task II the stimuli were presented in the form

1 2 3 4 5 6 7 8
* D A I N S C T E

for which the solution was to be reported as the word "DISTANCE" and the rule appeared in numerical form as follows: "1357 2468." The symbols or numbers above the letters were there to facilitate the employment of the rule. Other rules used in Task I were "The last pair of letters backward, then the first pair of letters forward, then the third pair of letters backward, and, finally, the second pair of letters forward" (in Task II the order given on the rule card was 87 12 65 34); "Each pair of letters in order, but the members of each pair are reversed" (in Task II the order on the rule card was 21 43 65 87); and, "The last pair forward, and the remainder in reverse order" (in Task II the order was 78 654321). Thus, in the transfer task, Ss in the associative facilitation group were required to employ another but similar version of a previously learned solution to a familiar situation.

Warm-up control. This group first learned a set of S-R pairs (D-E) that was unrelated to those in the transfer task (A-B). A question mark, a slash, a colon, and dollar sign

were used as discriminative stimuli and the solutions were learned as formulas comprised of number symbols (e.g., 876 12345). All stimuli and response-solutions in Task I were different from those in Task II. Thus, the transfer task for the warm-up control group required S to learn new solutions to new problem-solving situations.

Classical interference. The Task I list (A-E) learned by this group consisted of the same discriminative stimuli used in the solution of anagrams in Task II but they were paired with the formulas used in Task I by the warm-up control. Thus, the discriminative stimuli in both Task I and Task II for this group were identical but the response-solutions were completely different. In the transfer task S learned new solutions to a familiar problem-solving situation.

Associative interference. Task I, S-R pairs (A-B') learned by this group consisted of the discriminative stimuli and verbal rules used in the associative facilitation condition except that the rules were randomly re-paired with other stimuli. Thus, in Task II this group had to make new associations between learned solutions and familiar problem-solving situations. In addition, they were required to translate a verbal rule to a numerical rule to arrive at the solution.

Re-paired control. In Task I this group learned to pair the same discriminative stimuli and formulas required to solve the anagrams in Task II except that the pairing was rearranged. The experimental arrangement for this group (A-B) is described further in the introduction. As with the associative interference group, it was unnecessary for Ss to proceed through the translation step since the response-solutions were identically the same in both Task I and Task II.

Procedure

Each anagram was typed in capital letters in the center of 3 × 5 inch plain index cards with two spaces between each pair of letters. The cards were spiral-bound in booklets. The cards with solution rules were bound in separate booklets.

In Task I the booklets were placed in front of S on a table at which S and the experimenter were seated opposite each other. The S was instructed according to anagram instructions adapted from Ronning (1965). The overall procedure was, otherwise, similar to the anticipation method in paired-associate learning except that the trials were self-paced by S.

Briefly, *S* was presented an eight-letter anagram and was asked to spell the required nonsense word. In the first block of trials for Task I *S* was shown an anagram simultaneously with the corresponding rule for the Facilitation and Associative Interference groups but in numerical form for the Warm-up, Classical Interference, and Repaired Control groups. The rule indicated how the letters of the anagram were to be reordered. The *S*'s task was to call off the letters in the required order. He was given practice on a sample card until the anagram could be solved without the use of the rule card. Following the familiarization period he proceeded with the Task I training-test series. As he went through the training (Task I) deck he was instructed that he should try to recall the rule associated with the symbol (i.e., the discriminative stimulus). After eight training trials, two anagram problems for each of four different S-R combinations, *S* was tested on four new nonsense-word problems without knowledge of results, as a criterion for Task I. If he was unable to read off the four test words without error, the training series was again repeated. He was then tested on another, different, set of four test problems. The procedure was repeated until *S* reached the criterion of four perfect solutions when unscrambling the anagrams for the nonsense words in Task I.

A list composed of 16 eight-letter English nouns had been chosen for Task II from the Thorndike-Lorge (1944) word book. The procedure was to select a number at random between 1 and 208, the number of suitable pages in the Thorndike-Lorge norms. Beginning at the top of that page, that, and following pages if necessary, were perused until an eight-letter noun was found. When a word was found, the procedure was repeated. The nouns selected in this manner and used for anagrams in Task II were the following: distance, pleasure, anything, minister, evidence, stranger, addition, thousand, boundary, campaign, strategy, shoulder, business, position, division, language.

Task II was begun immediately following Task I with no break except to instruct *S* that he should now say the "eight-letter English word" to be deciphered from each of the new anagrams. The *S* was permitted to use pencil and paper if he thought it would help him. However, he was not permitted to write the rules. Time to solution was recorded by the experimenter with a maximum of 120 seconds allowed for each

anagram. If *S* had not reached solution within the allotted time, the rule was exposed for him to use. As in Task I, the appropriate rule was exposed for 5 seconds following a correct solution. The times to solution for the 16 nouns were blocked into four trials of four words each to facilitate the analysis of the data.

RESULTS

The mean number of trials to criterion for all groups on the initial task (Task I) was 3.5. The range of means for the treatment groups was 3.0-4.3. An analysis of variance of the data for the five groups indicated no significant differences among the means ($F = 1.99$, $df = 4/70$, $p > .05$).

The main transfer results for the treatment groups' performances on Task II are displayed in Figure 1. In

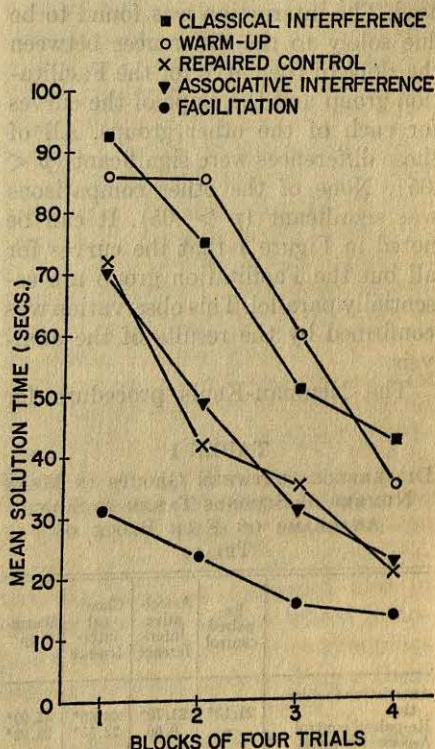


FIG. 1. Mean number of seconds taken by the five groups to solve the anagrams in Task II.

that figure are summarized the mean number of seconds taken by each group to solve each of the four blocks of four anagrams.

A mixed analysis of variance with five levels of the treatment variable (between) and four levels of trials (within) was performed on these data. The effect due to experimental treatments was significant ($F = 21.16$, $df = 4/70$, $p < .01$). In addition, the analysis yielded significant effects due to blocks of trials ($F = 80.39$, $df = 3/210$, $p < .01$), and to the interaction between treatments and blocks of trials ($F = 3.05$, $df = 12/210$, $p < .01$).

An analysis of differences in trends for the significant interaction term was made by the method of linear contrasts in which Scheffé's test was used. The interaction was found to be due solely to the differences between the slope of the curve for the Facilitation group and the slope of the curves for each of the other groups. All of these differences were significant ($p < .05$). None of the other comparisons was significant ($p > .05$). It can be noted in Figure 1 that the curves for all but the Facilitation group are essentially parallel. This observation was confirmed by the results of the analysis.

The Neuman-Keuls procedure for

TABLE 1

DIFFERENCES BETWEEN GROUPS IN MEAN
NUMBER OF SECONDS TAKEN TO SOLVE
ANAGRAMS ON EACH BLOCK OF
TRIALS

	Re- paired control	Associ- ative inter- ference	Classi- cal inter- ference	Warm- up
Associative facilitation	21.10*	21.70*	43.46*	45.02*
Re-paired control	—	0.60	22.53*	23.92*
Associative interference		—	21.93*	23.32*
Classical interference			—	1.39

* $p < .01$, Neuman-Keuls test.

making a posteriori tests was used to test the differences between experimental group means across all of the groups. The results of these analyses are presented in Table 1. There it may be seen that the performance of the Re-paired Control group did not differ significantly ($p > .05$) from that of the Associative Interference group. Neither did the performance of the Classical Interference group differ significantly ($p > .05$) from that of the Warm-up group. All other comparisons among treatments were significant ($p < .01$).

DISCUSSION

The evidence clearly supports the hypothesis that Task I experience with application of a solution rule transfers to benefit or hinder performance in a later problem-solving task. An interpretive analysis of these data permits the identification of three levels of interference-facilitation associated with transfer of solution rules in problem solving.

At the first level it is to be observed that the Facilitation group solved the anagrams in Task II more rapidly than did the remaining groups. In the first phase they had learned one form of the solution rules (response integration) and the stimuli to which these rules applied (hook-up or association). Only the translation to new forms of the rules had to be learned in the second task. It is perhaps all too obvious that associations of considerable strength must be assumed to have existed pre-experimentally between the verbal rule used in Task I and the number rule used in Task II by this group. However, to argue that variables other than mediational mechanisms were the main factors influencing the more rapid learning by this group compared to the other groups would

require an explanation in terms of general transfer principles. The variables that would necessarily account for transfer in such an analysis were controlled in other parts of the design. (See Jenkins, et al., 1965, for an earlier description on which the present explanation is based.) It should be noted that the Re-paired Control group had even more direct experience than did the Facilitation group with the specific stimuli and solution rules used in Task II, the exception being that the stimuli and solution rules were arranged differently in the two tasks. Accordingly, facilitation cannot be explained only in terms of learning experiences in the current experimental situation. Neither can the facilitation be attributed to a lack of equated interference possibilities or to nonspecific transfer. These factors have been controlled in the Warm-up and Classical Interference groups, both of which took longer to solve the anagrams in Task II than did the Facilitation group.

The Re-paired Control and the Associative Interference groups are at the second level of performance. These two groups had also achieved response integration through experience in the first task with the identical or similar forms of the rules to be used in the second task. However, the interchange of the pairing of these solution rules with discriminative stimuli imposed an additional learning requirement. Unlike the Facilitation group, the performance of these two groups was impaired by the requirement that new applications (or associations) had to be made. Although some interference resulted from these conditions, it does not appear to be as great, relatively speaking, as that observed in paired-associate learning when similar comparisons are made of the Facilitation

paradigm with either the Re-paired Control or Associative Interference paradigms.

Finally, at the third level of performance are to be found the Warm-up and Classical Interference groups. In these groups both the discriminative stimuli and solution rules, or only the solution rules, respectively, were changed from Task I to the transfer task. Thus, neither the solution rules or their applications, as learned in the first task, were applicable to solving the anagrams in the second task. It is this paradigm that has been employed most extensively in the water-jar and similar problem-solving tasks. The transfer task in these two conditions makes greater demands on *S* than do the other conditions. The *S* must integrate the new solution rule, link it with a new stimulus, and, possibly, learn to employ it to arrive at a correct solution. These last two groups may have benefited from experience in the first task as a result of nonspecific transfer (learning-to-learn) or they may have been hindered by interference from rules learned in the first task. However, the effects of these two factors cannot be evaluated in the present experiment due to the lack of an absolute control group.

The three levels described above differ in the degree of complexity of the learning process required in different stages of problem solving. Each level makes increasing demands on *S*, but the demands at one level are present at the next. This interpretative possibility is suggested also in the performance curves where it appears that the three major groupings begin at different points along a theoretical overall learning curve.

In general, the effects reported here on problem solving roughly parallel those found with verbal learning

tasks. Differences in results between the two learning situations are to be observed and to be expected because of differences in the processes required in the two kinds of tasks. In addition, the values of the variables in the present experiment could only be relative and thus differ markedly in value from those typically employed in paired-associate learning (Mandler & Heinemann, 1956).

REFERENCES

- DI VESTA, F. J., & WALLS, R. T. Transfer of object-function in problem solving. *American Educational Research Journal*, 1967, 4, 207-215.
- JENKINS, J. J., FOSS, D. J., & ODOM, P. B. Associative mediation in paired-associate learning with multiple controls. *Journal of Verbal Learning and Verbal Behavior*, 1965, 4, 141-147.
- LUCHINS, A. S. Mechanization in problem solving: The effect of Einstellung. *Psy-*

chological Monographs, 1942, 19(6, Whole No. 248).

- MANDLER, G., & HEINEMANN, S. H. Effect of over-learning of a verbal response on transfer of training. *Journal of Experimental Psychology*, 1956, 52, 39-46.
- REES, H. J., & ISRAEL, H. E. An investigation of the establishment and operation of mental sets. *Psychological Monographs*, 1935, 46(6, Whole No. 210).
- RONNING, R. R. Anagram solution times: A function of the "ruleout" factor. *Journal of Experimental Psychology*, 1965, 69, 35-39.
- SCHULZ, R. W. Problem solving behavior and transfer. *Harvard Educational Review*, 1960, 30, 61-77.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University Press, 1944.
- TWEDT, H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *Journal of Experimental Psychology*, 1959, 58, 111-116.
- UNDERWOOD, B. J. *Experimental psychology*. New York: Appleton - Century - Crofts, 1966.

(Received October 31, 1966)

TWO METHODS FOR ADAPTING SELF-INSTRUCTIONAL MATERIALS TO INDIVIDUAL DIFFERENCES¹

RALPH J. MELARAGNO

System Development Corporation, Santa Monica, California

2 methods for adapting self-instruction to individual differences (Branching and Prediction) were compared with each other and with a nonadaptive method (Linear). Branching used learner performance on the task for modifying the sequencing of instruction; Prediction used scores on 5 pretests for varying the sequence. Branching and Prediction procedures 1st were determined empirically, then the 3 methods were experimentally assessed in a computer-based laboratory. Bivariate analysis of training time and posttest score showed a significant difference between Branching and Linear groups. Univariate analysis of posttest score yielded no significant differences; univariate analysis of training time showed Branching superior to Prediction, and Prediction superior to Linear. Results support the use of procedures for adapting instructional materials to individual differences.

A continuing problem in the application of human learning research to instructional procedures is that of adapting to individual differences among learners. While it is recognized in the fields of education and training that individual differences in the learning process exist, little has been done about individuality because of the dearth of factual information derived from experimentation (Eckstrand, 1962).

One approach to human learning, programmed instruction (PI), has paid at least nominal attention to individual differences. Early forms of PI provided for individuality only in the learners' rate of progress through the instruction; a basic assumption was that the program was appropriate for all learners, and the only variation in performance would be in terms

of the rate variable. This form of PI has become known as "linear"; that is, all learners progress through a common program in a straight-line fashion.

A more recent form of PI has evolved through the introduction of the digital computer into the physical makeup of the teaching machine. Computer-based instruction uses a form of PI called "branching," permitting learners to branch at frequent times during training to the instruction supposed to be most appropriate to their individual needs. The basic branching system rests upon the computer's capability to keep track of learner performance data and to select the next sequence of instruction on the basis of prior performance. Thus, branching PI can accommodate individuality in amount of instruction and kind of instruction, as well as in rate of progress.

Recently, there have been proposals to use the preinstructional characteristics of the learner as criteria for the selection of appropriate types of instruction. Saltzman (1964) and Stolurow and Davis (1965) have indicated that entry behavior of the

¹ This paper is based on a doctoral dissertation presented to the Graduate School, University of Southern California. The research was conducted under the auspices of Contract N00014-66-C0081 between the Office of Naval Research, Washington, D. C., and System Development Corporation, Santa Monica, California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

learner be input to the machine and used for specifying the instruction appropriate for each learner.

Two distinct procedures, then, have been proposed for the modification of instruction as a function of the individual differences among learners. In one case, adaptation of a basic sequence is made on the basis of learner performance on prior instruction; in the second, the instructional sequence for any learner is determined by his unique preinstructional behavior. The present study was addressed to two questions: (a) Which of these procedures is more effective in adapting PI to individual learner differences? (b) Is either of them more effective than linear PI?

METHOD

Three procedures for presenting self-instructional materials were used in the study. The first, Branching, used data from learner performance on the instructional materials as a basis for adjusting subsequent instruction to individual differences. The second, Prediction, used data from learners' entry behavior for predicting the best sequences of instructional materials for each learner individually. The third, Linear, presented an identical instructional sequence to all learners.

The study was conducted in two phases. During the first phase, empirical trials of a self-instructional program in geometric inequalities were conducted with small groups of learners in order to determine strategies to be employed for the Branching and Prediction conditions. In the second phase, the relative effectiveness of the Branching, Prediction, and Linear conditions was experimentally investigated in a computer-based instructional laboratory.

Phase I

Subjects. Thirty-two volunteer subjects (Ss) were obtained from two high schools and were paid for their participation. The Ss were recruited from first-semester geometry classes, since students in these classes were naïve concerning the content of the program yet possessed the mathematical prerequisites.

Materials and apparatus. The program in geometric inequalities was prepared for and used in earlier investigations (Coulson, 1964; Coulson, Melaragno, & Silberman, 1965). It was made up of six units: The first five treated axioms, theorems, and postulates of inequalities and the sixth contained instruction on methods for preparing to prove theorems. At the end of each unit, Ss were tested briefly on the material in that unit, and then given an additional set of materials devoted to discussions of questions in the quiz.

The program was written in multiple-choice format. Each step in the program was presented on a separate page. Many of these steps called for no response but were only to be read. For the Phase I trials, the correct answer to each question was written on the back of the page containing the question.

Measuring instruments. The posttest was composed of 33 questions and, because of multiple answers, had a total possible score of 57; 29 points were possible for tasks directly related to the content of the program, and 28 points were possible for points related to transfer situations.

In order to obtain a wide sampling of relevant entry behavior, seven tests were selected as potential pretests. One test measured achievement in fundamentals of geometry; the other six were found by using Guilford's Structure of Intellect model (Guilford & Merrifield, 1960) as a means for determining intellectual abilities that should bear a relationship to the learning of geometry. The six Guilford tests used were: Problem Solving, Symbolic Grouping, Match Problems, Operations Sequence, Hidden Figures, and Symbolic Reasoning.

Method. The 32 Ss were first administered the seven pretests, then given the inequalities program. The first 12 Ss received the program individually, the next 10 in pairs, and the final 10 in two groups of five Ss each. The experimenter observed each S's progress through the program, responded to questions from Ss, and kept detailed records of kinds and specific locations of unusual performances to determine indications of the need for branching. Two distinct types of branching were determined: locations in the program where Ss could afford to skip past redundant instruction and locations where performance indicated the necessity for additional remedial instruction.

The branching structure that evolved from these empirical trials had the following characteristics: nine points in the

program where satisfactory performance allowed *S* to branch past redundant instruction and 14 locations where poor performance on quiz questions caused *S* to be branched to remedial instruction. The final version of the program contained 248 items in the main portion, 20 items devoted to end-of-unit quizzes and feedback to the quizzes, and 38 remedial items.

Results. After the empirical trials, the seven pretests were scored and related to *Ss'* performance on the program. The technique of multiple cutoff scores was found to be most effective in determining which *Ss* should branch at each of the 23 decision locations in the program. Two pretests showing no relationship to performance were eliminated: the Symbolic Reasoning and Match Problems tests.

Phase II

Subjects. To investigate experimentally the two questions of the study, 44 *Ss* were recruited from a different high school from those that supplied *Ss* for Phase I. The *Ss* were obtained from seven classes in the second semester of geometry, and were paid for their participation. In order to assure naiveté with respect to geometric inequalities, teachers of the seven classes postponed any treatment of inequalities until after the experiment was completed.

Treatment conditions. The *Ss* were randomly assigned to three treatment conditions: Linear ($N = 14$); Prediction ($N = 15$); and Branching ($N = 15$). The design of the experiment was a one-way, random effects, multiple analysis of variance, with the three conditions as independent variables and with scores on the posttest and training times as dependent variables.

Materials and apparatus. The program in geometric inequalities was used as the experimental vehicle. It was contained in looseleaf notebooks, each item on a separate page. Following each of the six units in the program, *Ss* received the quiz on the topics included in the unit and evaluated the adequacy of their answers.

The experiment was conducted in CLASS, a 20-station computer-based instructional laboratory at System Development Corporation in Santa Monica, California. Each station was provided with a notebook containing the program, a device for the *Ss'* responses to questions, scratch paper, and a pencil. To control for novelty of the environment, all three treatments were administered in CLASS.

Method. Prior to the study, *Ss* were administered the pretests in group settings. The *Ss* were brought to the laboratory daily by chartered bus at the end of normal school days for experimental runs of 50-75 minutes.

The Linear condition was run first, then the Branching, and finally the Prediction condition. The first time a particular treatment group came to the laboratory, *Ss* were read a standard set of directions explaining the operation of the equipment and procedures to be followed. Each *S* progressed at his own rate. The experimenter answered questions about the equipment, but not about the content of the program. When an *S* completed the program he remained at his station, but was not allowed to look over the program.

The posttest was administered to all *Ss* in each treatment condition, as a group. It was given the day following the completion of the program by all *Ss* in that group.

For the Linear condition, all 14 *Ss* received the same 285 items from the program. This included 248 main instruction items, 20 items of quizzes and feedback on quizzes, and 17 remedial instruction items.

The five pretests for each of the 15 *Ss* in the Prediction condition were pretested in the computer. At each of 23 decision points in the inequalities program, the computer examined these scores and directed each *S* to the next item that had been predicted as appropriate for him.

At each decision point in the program for the Branching condition, the computer examined certain prior performances of each of the 15 *Ss* in that group and determined whether or not to branch the *S*. Most branching decisions were made as a function of *Ss'* evaluations of their performance on end-of-unit quizzes. (Evaluations of quiz performances were ignored in the Linear and Prediction conditions.)

TABLE 1
POSTTEST SCORES AND TRAINING TIMES

	Linear	Prediction	Branching
Posttest scores			
<i>M</i>	37.4	37.3	41.1
<i>SD</i>	9.5	6.5	5.8
Training times (in minutes)			
<i>M</i>	195.1	168.6	147.1
<i>SD</i>	27.8	28.7	38.6

TABLE 2

VALUES OF F FOR COMPARISONS OF PAIRS OF BIVARIATE CENTROIDS

	F	df	
Linear vs. Prediction	3.057	2, 26	.10 > p > .05
Linear vs. Branching	10.593	2, 26	p < .001
Prediction vs. Branching	3.263	2, 27	.10 > p > .05

Results. The means and standard deviations for both dependent variables, by groups, are presented in Table 1. Box's test of homogeneity of bivariate dispersions (Cooley & Lohnes, 1962) yielded an F of .622, which was not significant. The multivariate analysis of variance was computed using the technique presented by Rao (1952); this yielded an F of 11.571 which, for 4 and 82 degrees of freedom, is significant beyond the .0005 level of confidence.

Since the hypothesis of homogeneity of dispersions could not be rejected and since the hypothesis of random sampling from a common population was rejected (by the multivariate analysis of variance), tests of the experimental questions were performed. The experimental questions were tested using Hotelling's T^2 -test for bivariate centroids which is an extension of Student's t -ratio for univariate samples (Rao, 1952). T^2 is evaluated for significance by transforming it to F ; the results are presented in Table 2.

In order to gain further information about the effects of the treatment conditions, the two dependent variables were also subjected to univariate analyses. For posttest scores, $F = 1.301$; $df = 2/41$; and $p > .10$. For training times, $F = 10.341$; $df = 2/41$; and $p < .005$. Since a significant difference was found among the three training times, t -tests were applied to evaluate differences between pairs of treatment means. The results are shown in Table 3.

TABLE 3

VALUES OF t FOR COMPARISONS OF MEAN TRAINING TIMES

	t	df	
Linear vs. Prediction	2.510	41	$p < .05$
Linear vs. Branching	4.545	41	$p < .01$
Prediction vs. Branching	2.071	41	$p < .05$

DISCUSSION

Inspection of the means in Table 1 shows a relatively clear ordering of the three treatment conditions. The Linear condition is the least effective; the Branching condition the most effective; and the Prediction condition falls in between. The significant difference in bivariate centroids between the Branching and Linear conditions implies that an adaptation procedure based upon S performance on the learning task is a useful strategy to follow. However, since the Prediction condition does not differ significantly from either the Branching or Linear conditions, it is not clear whether adaptation to individual differences is necessary and, if it is, what approach is to be preferred.

Some clarification of this situation is afforded by the results of the analysis of training times. It appears that when a self-instructional program produces essentially equivalent learning among S s regardless of individual differences, significant savings in training time can be achieved by attending to such individuality in learning. In addition, it seems that when individual differences are to be considered, a method using performance on the learning task (branching) is preferable to a method based on pre-training assessment (prediction).

These conclusions, and their implications for adapting to individual differences, must be evaluated in light of the restrictions in this study. Geometry is a unique subject matter because of the interrelationships among topics and the orderly development of instruction. Furthermore, geometry students are a restricted sample of a high school population: Geometry is an elective subject, chosen by students who have had previous success

in mathematics. It may be that the conclusions presented are limited to highly organized subject-matter areas and to learners who have achieved success in prior, related areas.

REFERENCES

- COOLEY, W. W., & LOHNES, P. R. *Multivariate procedures for the behavioral sciences*. New York: Wiley, 1962.
- COULSON, J. E. Research with a program in geometric inequalities. (TM-895/102/00.) Santa Monica: System Development Corporation, 1964.
- COULSON, J. E., MELARAGNO, R. J., & SILBERMAN, H. F. Nonprogram variables in the application of programmed instruction. (TM-2176/201/00.) Santa Monica: System Development Corporation, 1965.
- ECKSTRAND, G. A. Individuality in the learning process: Some issues and implications. *Psychological Record*, 1962, 12, 405-416.
- GUILFORD, J. P., & MERRIFIELD, P. R. The structure of intellect model: Its uses and implications. *Reports from the Psychological Laboratory, No. 24*. Los Angeles: University of Southern California, 1960.
- RAO, C. R. *Advanced statistical methods in biometric research*. New York: Wiley, 1952.
- SALTZMAN, I. J. Letter to the editor. *Contemporary Psychology*, 1964, 9, 238.
- STOLUROW, L. M., & DAVIS, D. Teaching machines and computer-based systems. In R. Glaser (Ed.), *Teaching machines and programmed learning*, Vol. II. Washington: National Education Association, 1965. Pp. 162-212.

(Received October 31, 1966)

TEACHER VERBAL CUES AND PUPIL PERFORMANCE ON A GROUP READING TEST¹

GEORGE S. LAMB

Western Washington State College

Verbal cues encouraging (a) rapid work, (b) accurate work, or (c) no specified cues were administered by 36 female teachers to their 2nd- and 3rd-grade classes during a group reading test. Effects of the cues were evaluated in terms of the class performances on the tests. Data were treated by analysis of variance and Scheffé confidence intervals at the .01 and .05 levels of significance respectively. Independent sources of variation were (a) treatments, (b) grade levels, (c) sex, and (d) reading achievement levels. Results indicated that girls were more responsive to teachers' verbal cues than boys.

What effects do teachers' verbal cues of encouragement or warning have upon the behavior of pupils? Teachers constantly offer a wide variety of verbal cues in their classrooms, which are intended to modify the behavior of pupils. The assumption that these cues really do affect pupil behavior is generally taken for granted. Yet this assumption has rarely been tested experimentally under ordinary classroom conditions.

Most of the experiments of verbal cues and reinforcements have been laboratory studies. They have generally involved an experimenter working with one subject at a time and have measured performance on simple physical tasks (Stevenson, 1965). The results of these studies have limited application to group instruction in the classroom (Jackson, 1965).

Classroom experiments in this area have investigated reinforcements, either oral (e. g., Auble & Mech, 1953; Hurlock, 1925; Van De Riet, 1964) or written (e. g., Anderson, White, & Wash, 1966; Forlano & Axelrod, 1937; Wallen, 1964). With the ex-

ception of the study by Page (1958) of teachers' written comments, these studies have dealt with pupils independent of one another instead of pupils as integral parts of a classroom with its own group interactions.

The present experiment was an investigation of verbal cues administered during the time that pupils were working on the tasks, rather than of verbal reinforcements administered at the conclusion of the tasks. Specifically, the study sought to answer the following question: Are the responses of second- and third-grade pupils on two subtests of the *New Developmental Reading Tests* (Bond, Ballou, & Hoyt, 1963a) affected by the teacher's verbal cues to work rapidly or to work accurately? The treatment cues were administered to classroom groups rather than to individual pupils.

One widely accepted principle of learning is that cues which arouse motivation toward the achievement of an educational objective will increase the effectiveness with which that objective is achieved (Wallen & Travers, 1963). If the treatment cues have a straightforward effect, we would expect that classes receiving speed cues would attempt more task items than classes receiving accuracy cues or no cues. We would also ex-

¹This study is based on a dissertation submitted to the Graduate School of the University of Minnesota in partial fulfillment of the requirements for a doctoral degree. The author is indebted to Theodore Clymer for his valuable assistance and advice in conducting this study.

pect that classes receiving accuracy cues would achieve a higher percentage of items correct than classes receiving speed cues or no cues.

METHOD

Subjects

The sample was 18 second-grade classes and 18 third-grade classes to which female student teachers had been previously assigned. The classes were located in central area and suburban schools of metropolitan Minneapolis and St. Paul.

Variables

The tasks were two subtests of the *New Developmental Reading Tests*, a group reading achievement test for second- and third-grade pupils. In the first subtest the pupils are to mark the one word among four which identifies a picture. In the second subtest the pupils are to mark a set of pictures according to specific written directions. Pupils attempt as many items as they can within the time limits of the test. For this study, it was important that the distributions of scores should not be compressed because pupils ran out of task items during the testing. The time limits for the subtests were, therefore, reduced from 10 to 4 minutes and from 15 to 9 minutes, respectively. The test yielded measures of seven dependent variables: (a) the number of items correct on Subtest I, (b) the number of items correct on Subtest II, (c) the adjusted score for guessing on Subtest I², (d) the number of items attempted on Subtest I, (e) the number of items attempted on Subtest II, (f) the ratio of the number of items correct to the number of items attempted on Subtest I, and (g) the ratio of the number of items correct to the number of items attempted on Subtest II.

Procedures

The classes were randomly assigned to one of three treatments: (a) speed, (b) accuracy, and (c) control. The female student teachers were the experimenters and received prior training on the administration of the tests and the treatments. The treatments were administered while the classes

Subtest I. Word Recognition

Time	Cue
0 min.	Begin. Let's see how quickly you can work today.
1 min.	You people are doing very well. I see that you are getting a lot done.
2 min.	My, you people are fast workers.
3 min.	It looks to me as if many of you will get most of your work finished.
4 min.	Stop.

Subtest II. Comprehending Specific Instructions

Time	Cue
0 min.	Begin. Let's see how quickly you can work today.
1 min.	You people are doing very well. I see that you are getting a lot done.
2 min.	My, you people are fast workers.
3 min.	It looks to me as if many of you will get most of your work finished.
5 min.	I am pleased to see how quickly everyone is doing his work.
7 min.	I am happy to see that you are getting so much done.
9 min.	Stop.

FIG. 1. Schedule of speed cues.

were taking the tests. For the speed treatment, the experimenters gave verbal clues encouraging rapid work. For the accuracy treatment, they gave verbal cues encouraging accurate work. For the control treatment, no verbal cues were given. The content and timing of the cues were specified by a script (see Figures 1 and 2). The timing of the tests and the cues was by stop watch.

Analyses

Since it was possible that the treatments would have different effects in the two grade levels or on boys and girls, the analysis compared the group means in these categories and their combinations as well as comparing the general treatment effects. Another possibility was that the treatments would have different effects upon groups of pupils of differing reading ability. In order to examine this latter effect, the third subtest of the *New Developmental Reading*

²The adjusted score is the number correct minus $\frac{1}{3}$ the number wrong, and is the standard method of scoring the first subtest.

Subtest I. Word Recognition

Time	Cue
0 min.	Begin. Let's see how carefully you can work today.
1 min.	You people are doing very well. I see many perfect papers.
2 min.	My, you people are careful workers.
3 min.	It looks to me as if many of you will get most of your work right.
4 min.	Stop.

Subtest II. Comprehending Specific Instructions

Time	Cue
0 min.	Begin. Let's see how carefully you can work today.
1 min.	You people are doing very well. I see many perfect papers.
2 min.	My, you people are careful workers.
3 min.	It looks to me as if many of you will get most of your work right.
5 min.	I am pleased to see how carefully everyone is doing his work.
7 min.	I am happy to see that you are doing this work correctly.
9 min.	Stop.

FIG. 2. Schedule of accuracy cues.

Tests (Bond, Balow, & Hoyt, 1963b), was administered to the classes prior to the administration of the treatments. The scores on this subtest were sorted into approximately equal groups of "high" and "low" for second-grade boys, second-grade girls, third-grade boys, and third-grade girls. Thus, the seven scores on the treatment tasks were analyzed according to four independent variables and their combinations: (a) treatment, (b) grade level, (c) sex, and (d) reading achievement level nested within grade by sex. The pooled means for these four classifications and their combinations were examined for trends. Differences among comparable means were tested by analysis of variance at the .01 level of significance, and, when appropriate, by Scheffé confidence intervals at the .05 level of significance.

RESULTS

The means of the variables which are normally used to score the tests,

TABLE 1

ABSOLUTE DIFFERENCE AMONG MEANS OF BOYS AND GIRLS BY TREATMENTS FOR THE NUMBER OF ITEMS ATTEMPTED ON SUBTEST II

Sex, treatment	(BA)	(BC)	(GS)	(GA)	$\frac{C}{C}$
Boys, speed (BS) $\bar{X} = 18.63$	0.44	0.29	1.61	0.63	0.72
Boys, accuracy (BA) $\bar{X} = 19.07$		0.73	1.17	1.07	0.28
Boys, control (BC) $\bar{X} = 18.34$			1.90	0.34	1.01
Girls, speed (GS) $\bar{X} = 20.24$				2.24*	0.89
Girls, accuracy (GA) $\bar{X} = 18.00$					1.35
Girls, control (GC) $\bar{X} = 19.35$					

Note.— $N = 12$ classes for all groups.

* $p < .05$.

i. e., the number of items correct and the adjusted score for guessing, did not occur in any consistent pattern, and the differences were not statistically significant. The mean scores for the variables involving the number of items attempted and the ratio of items correct to items attempted did occur in the orders predicted. The speed treatment group did attempt more items than the accuracy or control groups, and the accuracy treatment group did achieve a higher ratio of items correct to items attempted than the speed or control groups.

TABLE 2

POOLED MEANS AND F RATIOS OF BOYS AND GIRLS

Dependent variable	Pooled means		F ratio
	Boys	Girls	
Number correct			
Subtest I	21.74	24.35	30.35*
Subtest II	10.36	12.04	39.37*
Number correct minus one-third the number incorrect			
Subtest I	18.99	22.18	34.96*
Number attempted			
Subtest I	29.41	30.02	2.85
Subtest II	18.68	19.20	2.43
Ratio of number correct to the number attempted			
Subtest I	.73	.80	37.59*
Subtest II	.57	.63	13.35*

Note.— $N = 36$ classes.

* $p < .001$, 1/30 df.

However, the differences among the means were not statistically significant.

One interaction effect was statistically significant. This was the Treatment $\times$ Sex interaction for the number of items attempted on the second subtest. The differences among the pooled means associated with this interaction are in Table 1. An analysis of these differences using Scheffé confidence intervals showed that girls attempted significantly more items under speed cues than under accuracy cues.

Girls consistently scored higher than boys on all seven variables, as shown in Table 2. The differences were statistically significant except on the two variables for the number of items attempted.

CONCLUSIONS AND DISCUSSION

The complete data (Lamb, 1965) are too numerous for this report, and must be interpreted with caution. It should be kept in mind that the total treatment period was 13 minutes and that the sample size was small.

Within those limitations, it would appear that the mean scores of two typical measures of achievement on a primary reading test are not significantly affected by consistent cues to work rapidly or carefully. This stability of result should be encouraging to the authors, publishers, and users of such tests.

The results indicate that girls may be affected significantly by speed and accuracy cues in the number of items they attempt. The fact that this effect was apparent on the second subtest and not on the first might be because the treatments had been in effect longer at the time of the second subtest, or it might be due to the nature of the tasks in the second subtest. Since boys were not affected in

this way, there is the implication that girls are more responsive to a female teacher's verbal cues than are boys.

The clear superiority of girls over boys was less pronounced on the measures of the number of items attempted. This is not easily explained. It may be that the dominant treatment on girls was the accuracy cues, and that these had the effect of reducing the number of items they attempted. Another possibility is that the superior achievement of girls on reading tests is due more to the accuracy with which they deal with the items than with the number of items they attempt.

REFERENCES

- ANDERSON, H. E., WHITE, W. F., & WASH, J. A. Generalized effects of praise and reproof. *Journal of Educational Psychology*, 1966, **57**, 169-173.
- AUBLE, D., & MECH, V. E. Partial verbal reinforcements related to distributed practice in a classroom situation. *Journal of Psychology*, 1953, **38**, 165-186.
- BOND, G. L., BALOW, E. B., & HOYT, C. J. *New developmental reading tests*, Form U-II. (Part I, Word recognition, Part III, Comprehending specific instructions.) Chicago: Lyons & Carnahan, 1963. (a)
- BOND, G. L., BALOW, E. G. & HOYT, C. J. *New developmental reading tests*, Form U-II. (Part II, Comprehending significant ideas.) Chicago: Lyons & Carnahan, 1963. (b)
- FORLANO, G., & AXELROD, H. C. The effect of repeated praise or blame on the performance of introverts and extroverts. *Journal of Educational Psychology*, 1937, **28**, 92-100.
- HURLOCK, E. B. An evaluation of certain incentives used in school work. *Journal of Educational Psychology*, 1925, **16**, 145-159.
- JACKSON, P. W. The way teaching is. *WEA Journal*, March 1965, **54**, 10-12, 16.
- LAMB, G. S. The effects of verbal cues upon second and third grade pupils' responses on a group reading test. Unpublished doctoral dissertation, University of Minnesota, 1965.
- PAGE, E. B. Teacher comments and student performance. *Journal of Educational Psychology*, 1958, **49**, 173-181.

- STEVENSON, H. W. Social reinforcement of children's behavior. In L. P. Lipsitt & C. C. Spiker (Ed.), *Advances in child development and behavior*. Vol. 2. New York: Academic Press, 1965. Pp. 97-126.
- VAN DE REIT, H. Effects of praise and reproof on paired-associate learning in educationally retarded children. *Journal of Educational Psychology*, 1964, **55**, 139-143.
- WALLEN, C. J. Teacher, individual, and group issued incentives and pupil performance: A nineteen classroom experiment in motivation. *Journal of Educational Research*, 1964, **57**, 413-416.
- WALLEN, N. E. & TRAVERS, R. M. W. Analysis and investigation of teaching methods. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963. Pp. 495-497.

(Received November 7, 1966)

ATTENTIONAL PROCESS IN READING: THE EFFECT OF PICTURES ON THE ACQUISITION OF READING RESPONSES

S. JAY SAMUELS¹

University of Minnesota

Pictures may be used as prompts when the reader cannot read a word in the text, but pictures may miscue and may divert attention from printed words. In Experiment I, 30 randomly assigned pre-1st graders learned to read 4 words with no picture, a simple picture, or a complex picture present. During acquisition trials, when pictures were present, the simple and complex picture groups made more correct responses ($p < .01$). During test trials, with no pictures present, the no-picture group excelled ($p < .01$). In Experiment II, 26 matched pairs of 1st graders were given classroom reading instruction under a no-picture or picture condition. The results disclosed that poor readers with no picture present learned more words ($p < .01$). Among better readers the difference was not significant. The results are discussed in terms of attentional processes.

Psychologists have long been aware of the central role of attentional processes in learning. Pavlov and his associates found that in order to classically condition animals distracting stimuli which competed for the animal's attention had to be eliminated. To accomplish this, they worked in a specially constructed soundproofed building with partitions separating experimenter and animal. The mere presence of the experimenter or the sound of footsteps in the experimental situation seriously prolonged the conditioning procedure (Osgood, p. 311, 1953). Distracting background stimuli not only interfere with the learning of animals but with the learning of humans as well. The ability of the individual to withhold attention selectively from irrelevant and distracting background stimulation seems to be implicated in reading disability, according to Santostefano, Rutledge,

and Randall (1965). It appears that when distracting stimuli are present the performance of the underachiever undergoes greater disruption than does the performance of the more capable student (Baker & Madell, 1965; Silverman, Davids, & Andrews, 1963).

The purpose of the present study was to test the hypothesis that when pictures and words are presented together, the pictures would function as distracting stimuli and interfere with the acquisition of reading responses. Pictures may be used as prompts when the reader cannot read a word in the text, but pictures may miscue and divert attention from the critical task of attending to the printed words. In order to test this hypothesis two experiments were conducted. Experiment I was designed as a laboratory study to test the effect of pictures on naïve subjects (Ss) under conditions unlike those found in classrooms. Since findings derived from laboratory settings are often subject to the criticism that they could not be replicated in less artificial situations, Experiment II was designed to test the ef-

¹ The author would like to express his appreciation to Joseph Jenkins for his help in data collection and data analysis for Experiment I, and to Edwin Myers for his help in data collection and data analysis in Experiment II.

fect of pictures on less naïve Ss using a procedure which was similar to that used in actual classrooms.

EXPERIMENT I

Method

Subjects. Thirty children who had kindergarten experience and who were enrolled in a prefirst grade summer program were randomly assigned to one of three experimental treatments. The Ss were pretested to assure that no one was able to read the words used in the experiment.

Design. A simple, randomized design was used. Ten Ss were randomly assigned to the no-picture, 10 to the simple-picture, and 10 to the complex-picture condition.

Materials. The pretest materials consisted of four 5 × 8 inch index cards with either "boy," "bed," "man," or "car" typed on them. There was one word typed on each card.

For the warm-up trials, four novel, non-sense figures that had been created for a study on reading were used (Jeffrey & Samuels, in press). The figures were highly discriminable from each other and sufficiently dissimilar from letters in the Roman alphabet to make generalization from the other unlikely. The figures were placed on 5 × 8 inch index cards, one to a card. The Ss had to learn to associate a number with each figure. The numbers were 1, 2, 3, and 4.

For the experiment proper, a primary typewriter was used to type the words, boy, bed, man, and car, at the bottom of 5 × 8 inch index cards, one word to a card. The same words were presented in the no-picture, simple-picture, and complex-picture conditions.

For learning trials in the no-picture condition, there was a word at the bottom of each card but no picture was present.

For learning trials in the simple-picture condition, the words were typed at the bottom of each card as in the no-picture condition. Above each word was a simple black and white drawing portraying the object that the word symbolized. Previous work with the pictures indicated they could reliably elicit the same verbal response as was typed at the bottom of the card.

For learning trials in the complex-picture condition, the words were typed at the bottom as in the no-picture condition. Above each word was a colorful picture which had been cut out of a reading primer. The

pictures were complex in that they pictorially represented more than the word which was at the bottom of the card. For example, the picture attached to the card with the word "boy" at the bottom showed a boy holding his dog. The boy was pointing to a horse in the background.

The cards used for the test trials were the same for all conditions. At the bottom of the test cards the four words, boy, bed, man, and car, were typed in lower case with the primary typewriter, one word to a card. There were no pictures on any of the test cards.

Procedure

The experimenter worked individually with the Ss during all phases of the procedure. A pretest was given to each S. The S was told, "Today, we are going to play a game. In this game we are going to learn some words. First, let us see if you already know what the names of the words are." The four words were shown to the S. If he was able to read any of the words, he was eliminated.

Following the pretest a warm-up was given to each S to acquaint him with the nature of the learning task. The S was told, "Before we learn the new words, let us practice on some numbers. I will show you a card with a funny-looking number on it and I want you to tell me what the number is. If you don't know the number's name I will tell you what it is. You should try to tell me what the number is before I tell you. Do you understand what we are to do? All right? Then, what do you do when I show you a card with a number on it?" Each card was shown to S for an approximate 4-second interval. At the end of the anticipation interval S was told the number. The cards were presented in three random orders. Each S was given six warm-up trials.

When the warm-up was over, S was given the first learning trial. He was told, "All right, now let us see how we can learn the new words. I will show you a card with a word on it and I want you to tell me what the word's name is. If you don't know the word's name I will tell you. You should try to tell me the name before I tell you. Do you understand?" Each card was shown to the S for 4 seconds. Then he was told the name. The S was scored for a correct response if he said the appropriate word before feedback was given.

Following the first acquisition trial, the

first test trial was given. The test card was presented for 4 seconds and *S* had to give the correct response during this interval. No feedback was given during test trials. Acquisition trials and test trials were alternated. Each *S* received 10 learning trials and 10 test trials. The stimuli for all phases of the procedure were presented in three random orders.

Results

Separate analyses were computed for responses during acquisition trials and for responses during test trials. During acquisition trials pictures were present as incidental cues for the picture conditions. During test trials pictures were not present.

On the acquisition trials, as seen in Table 1, the mean number of correct responses given for the no-picture condition was 25.30; for the complex picture condition it was 36.90; and for the simple-picture condition it was 39.40. Comparing the simple-picture to the no-picture condition during acquisition, *Ss* in the simple-picture condition gave significantly more correct responses ($t = 9.02$, $df = 18$, $p < .01$). Comparing the complex-picture to the no-picture condition, *Ss* in the complex-picture condition gave significantly more correct responses ($t = 7.42$, $df = 18$, $p < .01$).

On the test trials, where incidental cues were absent for all conditions the no-picture group excelled. As seen in Table 2, the mean number of correct responses given on the tests by *Ss* in the no-picture condition was 19.20; for *Ss* in the simple-picture condition it

TABLE 2
MEANS AND STANDARD DEVIATIONS FOR
TREATMENTS ON TEST TRIALS

Treatment	<i>M</i>	<i>SD</i>
No-picture	19.20	7.93
Simple-picture	11.30	5.79
Complex-picture	11.60	4.93

was 11.30; and for *Ss* in the complex-picture condition it was 11.60. Comparing the simple-picture to the no-picture condition on the test trials, *Ss* in the no-picture condition recognized significantly more words ($t = 4.02$, $df = 18$, $p < .01$). Comparing the complex-picture to the no-picture condition, *Ss* in the no-picture condition recognized significantly more words ($t = 3.87$, $df = 18$, $p < .01$).

Discussion

The purpose of Experiment I was to test the hypothesis that when related pictures and words are presented together, the presence of pictures would retard the acquisition of reading responses. This experiment used naive *Ss* under conditions unlike those found in classrooms. The results disclosed that during the 10 acquisition trials, when pictures were available as incidental cues for appropriate verbal responses for *Ss* in the picture conditions, *Ss* in these conditions gave significantly more correct responses than did *Ss* in the no-picture condition. On the 10 critical test trials, when pictures were not available as incidental cues, *Ss* in the no-picture condition gave significantly more correct responses. Acquisition and test trials were purposely alternated so that *Ss* in the picture conditions would be aware that the printed words were important stimuli. Despite the alternation of acquisition and test trials, *Ss* in the picture con-

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR
TREATMENTS ON ACQUISITION TRIALS

Treatment	<i>M</i>	<i>SD</i>
No-picture	25.30	7.23
Simple-picture	39.40	1.26
Complex-picture	36.90	6.34

ditions tended to use pictures rather than words as cues. It would appear that pictures functioned as distracting stimuli in that they drew attention away from the printed words. A similar finding has been reported by Underwood (1963) and Samuels and Jeffrey (1966). They report that in paired-associate learning, under certain conditions, an *S* may extract from the stimulus complex an incidental cue to which the response gets attached. When the stimulus complex is presented, the *S* may give the correct response, but for the wrong reasons. For example, when presented h-o-r-s-e, the child says "horse." He may have attached the response to the letter "h," however. When the stimulus h-o-u-s-e is presented, he says "horse" since he attends only to the one letter.

Subsequent to the completion of Experiment I, two doctoral dissertations were conducted which tested the effect of pictures on reading acquisition. Despite methodological differences, these studies lend additional support to the finding that the presence of pictures retards reading acquisition. Using kindergarteners, Braun (1967) found significant differences in reading acquisition favoring the no-picture group on seven of eight comparisons. Harris (1967), who used kindergarteners from a low socioeconomic background, found significant differences in acquisition on four of eight comparisons, favoring the no-picture group. Harris attributed his failure to find significance on more comparisons to the generally low level of learning for all his *Ss* regardless of experimental condition. In both studies, all comparisons which did not reach significance were in the predicted direction.

EXPERIMENT II

Method

Subjects. Fifty-two students from a Minneapolis public school with 7 months first-grade experience were used as *Ss*.

Design. A treatment by levels design was used. The *Ss* were divided into two matched groups. There were 26 *Ss* in the picture and 26 *Ss* in the no-picture condition. Half the *Ss* in each condition were designated as above and the other half as below the median based on pretest scores. The same test was used as pretest and posttest for all *Ss*, regardless of condition.

Materials. A pretest and posttest were used in this study. Both tests were exactly the same. The test was constructed by typing in large type each of the 50 different words used in the story "Fun at Blue Lake" on 3 × 5 inch index cards, one word to a card.

The reading material for the no-picture and picture condition was the same. A story called "Fun at Blue Lake" was written. It was 106 words long and contained 50 different words. The words were typed in large type and the story was mimeographed. A book was made for each *S*. The story, "Fun at Blue Lake," was pasted on the right face of the book for both conditions. The *Ss* in the picture condition had a picture from a reading primer pasted on the left face of the book. The picture showed a cabin in the woods with a lake in the foreground. At the lakeshore was a family and their dog. In the no-picture condition the left face of the book was blank. When the books were opened *Ss* in the picture condition saw the text on the right and a picture on the left. The *Ss* in the no-picture condition saw a blank page on the left and the text on the right.

Procedure

Several days before the experiment proper was run, *Ss* were pretested on the 50 words used in the story. The pretest consisted of showing each of the 50 words used in the story to *S*. The words were exposed one at a time to *S*, allowing 10 seconds for a response. The *Ss* were then matched on pretest scores and randomly assigned to either the picture or no-picture condition.

Reading instruction was given to both conditions simultaneously. The groups were separated in the room so that *Ss* in the no-picture condition were unable to see the pictures of the *Ss* in the picture condition. Reading instruction was given to small

groups. At no time were there more than eight Ss in the room. The instructional procedure paralleled that used in typical classrooms. Instruction consisted of motivating and building background for the story, reading for a purpose, silent reading, and then oral reading. The Ss were instructed to raise their hands during silent reading if they were unable to read any word. The experimenter went about whispering the words to the children who requested help. The experimenter was careful to give help to students in both groups and both groups were given opportunities to read aloud.

Immediately following reading instruction, the Ss were given the posttest. They were tested by four assistants on the 50 words used in the story in a similar manner to that which was used in the pretest.

Results

Two comparisons were made to test differences in word recognition on the posttest. In the first comparison, as seen in Table 3, above-median Ss in the picture condition recognized a mean of 43.15 words while above-median Ss in the no-picture condition recognized a mean of 42.08 words. This difference was not significant. In the second comparison, below-median Ss in the picture condition recognized a mean of 23.69 words, while below-median Ss in the no-picture condition recognized a mean of 26.23 words. This difference was significant ($t = 2.73$, $df = 12$, $p < .01$, one-tailed test).

Discussion

Experiment II was designed to test the effect of pictures on reading acquisition using a procedure which was similar to that used in classrooms. The Ss in this experiment had 7 months of formal reading instruction. No significant difference was found in reading acquisition between the picture and no-picture condition among the better readers. Among the poorer readers, Ss in the no-picture condition

TABLE 3
WORD RECOGNITION SCORES ON POSTTEST
FOR ABOVE AND BELOW MEDIAN SUBJECTS
FOR TREATMENTS

Reading ability	M	SD
Below median		
No-picture treatment	26.23	8.48
Picture treatment	23.69	7.69
Above median		
No-picture treatment	42.08	4.57
Picture treatment	43.15	6.05

learned to read significantly more words than did Ss in the picture condition. The results of Experiment II support the findings of Silverman, Davids, and Andrews (1963) and Baker and Madell (1965) in that performance of the less capable student was affected more by distracting stimuli than was the performance of the more capable student.

Several questions are raised by the two experiments. Considering the effect which pictures had on reading acquisition of naïve and less capable students, one may wonder if it is good practice to put pictures in reading primers. Another question left unanswered is how the more capable student uses pictures when they are available. It would seem advisable to continue the investigation of the role of pictures in readers in terms of motivation, student attitude towards reading, and attentional processes.

REFERENCES

- BAKER, R. W., & MADELL, T. O. A continued investigation of susceptibility to distraction in academically underachieving and achieving male college students. *Journal of Educational Psychology*, 1965, 56, 254-258.
- BRAUN, C. The efficacy of selected stimulus modalities in acquisition and retention of sex-typed textual responses of kindergarten children. Unpublished doctoral dissertation, University of Minnesota, 1967.

- HARRIS, L. A. A study of the rate of acquisition and retention of interest-loaded words by low socio-economic kindergarten children. Unpublished doctoral dissertation, University of Minnesota, 1967.
- JEFFREY, W. E., & SAMUELS, S. J. The effect of method of reading training on initial learning and transfer. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 354-358.
- OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford University Press, 1953.
- SAMUELS, S. J., & JEFFREY, W. E. Discriminability of words and letter cues used in learning to read. *Journal of Educational Psychology*, 1966, 57, 337-340.
- SANTOSTEFANO, S., RUTLEDGE, L., & RANDALL, D. Cognitive styles and reading disability. *Psychology in the Schools*, 1965, 2, 57-62.
- SILVERMAN, M., DAVIDS, A., & ANDREWS, J. M. Powers of attention and academic achievement. *Perceptual and Motor Skills*, 1963, 17, 243-249.
- UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning, problems and processes*. New York: McGraw-Hill, 1963. Pp. 33-48.

(Received November 14, 1966)

PREDICTION OF STUDENT ACCOMPLISHMENT IN COLLEGE

JAMES M. RICHARDS, JR., JOHN L. HOLLAND, AND SANDRA W. LUTZ
American College Testing Program, Iowa City, Iowa

In samples with a broad range of talent, the academic and non-academic achievements of college students were predicted. Criteria included college grades, 12 scales designed to measure notable extra-classroom accomplishment in college, and 1 scale to assess recognition for academic accomplishment. Predictors included scores on ACT tests, high school grades, and 6 scales measuring nonacademic accomplishment in high school. Results indicate that nonacademic accomplishment can be assessed with moderate reliability, that both academic and nonacademic accomplishment can be predicted to a useful degree, and that nonacademic accomplishment is largely independent of academic potential and achievement.

The present study aims to predict student achievement in college from a comprehensive assessment of student achievement and potential in high school. Previous studies designed to predict academic and extracurricular achievement in college for students of superior scholastic aptitude (Holland, 1959, 1960, 1961; Holland & Astin, 1962; Holland & Nichols, 1964; Nichols & Holland, 1963) are extended by this study, which is similar to them in its goals and longitudinal method. It differs from them, however, in that predictions are made for students with a *broad range* of academic potential.

The present study is also related to many other investigations of similar problems. Among these problems are the relationship between academic potential and originality, the description of creative persons, the development of criteria of creative performance, and the prediction of adult accomplishment. Researchers who have worked on such problems include: Astin (1962); Barron (1963); Buel (1965); Chambers (1964); Cicirelli (1965); Flescher (1963); Getzels and Jackson (1962); Gough, Hall, and Harris (1963); Guilford (1964); Hoyt (1966); Locke (1963); MacKin-

non (1960); Mann (1958); Price, Taylor, Richards, and Jacobsen (1964); Skager, Schultz, and Klein (1965); Sprecher (1959); Taylor, Smith, and Ghiselin (1963); Thorndike and Hagen (1959); Torrance (1963); and Wallach and Kogan (1965).

The rationale for this study is that typical measures used in the selection of college students—tests of academic potential and high school grades—concentrate on only one dimension of talent and ignore other important dimensions (Holland & Richards, 1965). Accordingly, if we want to find college students who will do outstanding things outside the classroom and in later life, we need a record of student achievements outside the classroom in high school. The present study examines the predictive validity of one such record of student achievement.

METHOD

Predictors

The predictive variables included the following measures:

1. ACT tests. The test battery, a college admissions test administered nationally, yields the following subtest scores: English, mathematics, social studies, and natural science. Each score is converted to a common

scale with a mean of approximately 20 and a standard deviation of about 5 for college-bound high school seniors. The reliabilities of the ACT tests (American College Testing Program, 1965), the high correlations between the ACT battery and other similar measures (Ellis, 1962), and the similar relationship of the ACT battery and of similar measures to college grades (Munday, 1965) all indicate that the ACT battery is a typical measure of academic potential. Therefore, we would not expect markedly different results in the present study if we had used some other academic test or test battery.

2. High school grades. As a regular part of the ACT procedure, persons taking the ACT battery report the grades they have received in high school courses in four areas: English, mathematics, social studies, and natural science. Research by Davidsen (1963) indicates that in a large sample such self-reported grades correspond closely to the high school transcripts. A reanalysis of Davidsen's data by the present authors yielded a correlation of .92 between student-reported and school-reported grades. The measure used in the present study is the overall average on a 4-point scale ($A = 4$, $B = 3$, etc.) of all grades reported. In another study by Hoyt (1964) the predictive efficiency of average self-reported grades equaled the predictive efficiency of the student's rank in the high school class obtained from his transcript.

3. Extracurricular achievement record. Checklists of extracurricular accomplishment for the high school years were used to obtain scores in the following areas: art, music, literature, dramatic arts, leadership, and science (Holland & Nichols, 1964). Items ranged from common and less important accomplishments to rare and more important ones. For example, science items included accomplishments such as: did an independent scientific experiment; won a prize or award of any kind for scientific work or study; had scientific paper published in a scientific journal. The remaining scales consisted of similar items planned to assess a broad range of achievement. The score on each scale is simply the number of accomplishments the student has attained.

The achievement record was obtained as part of the American College Survey. The Survey booklet contains several sections designed to elicit information about a student's aspirations, achievements, attitudes, interests, potentials, values, and background (Abe, Holland, Lutz, & Richards, 1965). In

the American College Survey sample, the reliabilities (K-R 20) of the achievement scales ranged from .72 to .84 for men and from .65 to .81 for women.

Student Sample

The student sample was obtained from a follow-up of students who participated in the American College Survey (Abe et al., 1965). In the original study, a comprehensive assessment was administered to 12,432 college freshmen in 31 institutions of higher education during the months of April or May of 1964. The sample for the present study is restricted to the 7208 students at 22 of the 29 colleges participating in the follow-up study who also took the American College Testing battery in the academic year 1962-63 as part of their application for admission to college. The record of college accomplishments for these students was obtained in the spring of 1965 at the end of their *sophomore* year in college.

In September of 1964, a second study involving the American College Survey was conducted in which the same comprehensive survey was administered to 5668 entering freshmen at six colleges.¹ This second sample of 2483 is also restricted to the freshmen in the larger group who took the American College Testing battery as part of their application for admission to college. The follow-up data for these students also were collected in the spring of 1965 at the end of their *freshman* year in college.

Each college was responsible for the administration of the follow-up questionnaire. Several techniques were used to contact students: Some colleges had students fill out the questionnaire in English classes, convocations, or other group sessions; other colleges polled their students by mail. Complete follow-up data were obtained for 2792

¹ Tables showing the colleges for the two samples and the details of certain statistical analyses discussed later (t tests comparing students with and without follow-up data, complete intercorrelation matrices, and the correlation between academic predictors and all criteria at individual colleges) have been deposited with the American Documentation Institute. Order Document No. 9572 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D.C. 20450. Remit in advance \$1.25 for photocopies or \$1.25 for microfilm and make checks payable to: Chief, Photoduplication Service, Library of Congress.

sophomore students (1373 men and 1419 women) and 1095 freshman students (503 men and 592 women). Follow-up data were thus obtained for 39% of the sophomores and 44% of the freshmen. Students with missing follow-up data include both students who left college and students still enrolled in college who failed to complete the follow-up questionnaire.

Because this is a low return rate, it is important to know what biases there may be in the sample with follow-up data. Accordingly, *t* tests were computed between students with and without follow-up data on each of the predictor variables in each of the groups. While each of these *t* tests is not completely independent of every other test (some of the variables are correlated to a substantial degree), for the purposes of this study, any error introduced is conservative since it is more likely that a number of significant differences will be found between students with and without follow-up data.

The primary trend in the results is for students with missing follow-up data to have significantly lower ACT scores and high school grades. This is to be expected, of course, since this group includes students who left college because of academic failure. However, because the *N*s in this study are very large, a small absolute difference can be highly significant. The actual differences on ACT scores and high school grades between students with and without follow-up data are not large relative to the standard deviations of these variables. On the extracurricular achievement scales, only a few differences are significant, and these fall into no consistent pattern. It appears, therefore, that although there are some significant differences between students with and without follow-up data, it is unlikely that the results of this study are seriously distorted by these differences because virtually a full range of accomplishment is present in the groups with follow-up data.

To summarize, because the colleges used such diverse means of administering the survey and because there are significant differences between students with and without follow-up data, our samples may not be a precise representation of the college populations included. Nevertheless, our samples do represent a broad range of students from diverse institutions. Because most earlier studies of this problem were based on a narrow range of talent, the present samples permit a more definitive examination of the relationships in question.

Criteria of Achievement

The criterion variables included the following measures:

1. College grades. Each student reported his grade average for his last college term by checking one of the following alternatives: D or lower, D+, C, C+, B, B+, A or A+. Scores from 1 to 7 were assigned to these alternatives so that a high score indicates high grades.

2. Nonclassroom achievement record. A checklist of nonacademic accomplishments was developed to measure achievement in the following areas: leadership, social participation, art, social service, science, business, humanities, religious service, music, writing, social science, and speech and drama. A simple scale was also developed to determine public recognition for academic attainment in college. Each scale is, in a sense, a criterion or standard of accomplishment in an important area of human endeavor. Students with high scores on one or more scales are assumed to have attained a high level of accomplishment which required complex skills, long term persistence, or originality, and which generally received public recognition. A detailed account of the rationale, development, and statistical characteristics of these scales is presented elsewhere (Richards, Holland, & Lutz, in press).

Each scale includes ten items, except the Recognition for Academic Accomplishment Scale, which has five items. In responding to the items, the student marks "yes" for those accomplishments which he has achieved during college and "no" for those which he has not achieved. The score on each scale is simply the number of "yes" responses.

Items range from common and less important accomplishments to rare and more important ones. For example, leadership accomplishments included: elected to one or more student offices, active member of four or more student groups, served on a student-faculty committee. Music accomplishments included: composed or arranged music which was publicly performed, publicly performed on two or more musical instruments (including voice) which do not belong to the same family of instruments, attained a first-division rating in a state or regional solo music contest. The remaining scales consisted of similar items with content appropriate to the various areas of achievement. In general, the accomplishments involve public action or recognition,

TABLE 1
MEANS AND STANDARD DEVIATIONS ON COLLEGE ACHIEVEMENT
SCALES FOR THE STUDENT SAMPLES

Scale	Men				Women			
	Freshmen		Sophomores		Freshmen		Sophomores	
	M	SD	M	SD	M	SD	M	SD
Scientific achievement	.18	.68	.32	.86	.07	.32	.10	.39
Leadership achievement	.63	1.30	.88	1.53	.72	1.33	1.40	1.83
Speech and dramatic achievement	.31	.81	.30	.90	.26	.75	.36	.93
Artistic achievement	.38	.87	.48	1.07	.67	1.20	.85	1.34
Writing achievement	.31	.73	.27	.73	.39	.75	.50	1.01
Musical achievement	.16	.61	.21	.73	.14	.52	.27	.73
Social participation	.80	1.35	.90	1.39	.77	1.30	1.03	1.34
Social service achievement	.55	1.07	.70	1.21	.83	1.24	1.20	1.40
Business achievement	.54	.91	.68	1.00	.22	.51	.34	.65
Humanistic-cultural achievement	.94	1.21	1.04	1.33	1.23	1.42	1.45	1.48
Religious service	.73	1.48	1.34	2.20	1.30	2.06	1.98	2.41
Social science achievement	.24	.57	.33	.70	.27	.54	.32	.60
Recognition for academic accomplishment	.14	.46	.36	.69	.14	.40	.44	.81

Note.—For freshmen men, $N = 503$; for sophomore men, $N = 1373$; for freshmen women, $N = 592$; for sophomore women, $N = 1419$.

so that, in principle, they could be verified. It was assumed that such possibility of verification would lessen student exaggeration and allow a comparison of student self-reports with public records.

The nonclassroom college achievement scales were administered as a part of a comprehensive follow-up of the American College Survey (Abe et al., 1965). The follow-up questionnaire elicited information about a college student's achievements, aspirations, self-concept, satisfactions, and attitudes.

RESULTS

The means and standard deviations of the college-achievement scales for the various samples are summarized in Table 1. The distributions of the nonacademic accomplishments are highly skewed and almost dichotomous, and the standard deviations are larger than the means.² This

² The skewness of such distributions has had little effect in previous studies, however, on Pearson correlations involving similar variables (Holland & Richards, 1965). It is possible that the results of this study are

skewness occurs because each scale contains accomplishments that are rare among college students. (The modal number of accomplishments on most scales is zero.) Differences among the areas of accomplishment probably reflect differences both in the level of accomplishment represented by the various items and in the opportunity for various kinds of achievement in college.

As a next step, correlations were computed among all of the variables, both predictor and criterion.³ In general, there are: (a) moderate correlations among measures of academic potential and performance, (b) moderate correlations among nonclass-

distorted by the use of skewed, almost dichotomous variables in the multiple regression analyses, although the consistency and meaningfulness of the findings suggest that such distortion is unlikely.

³ Computations for this study were carried out at Measurement Research Center, University of Iowa, and at the University of Utah Computer Center.

room achievements in the same or closely related areas, (c) low to moderate correlations among nonclassroom achievements in areas which are not closely related, and (d) low relationships between nonclassroom achievements and measures of academic potential and performance. The relationships are consistent with what previous investigators have found (Holland, 1959, 1960, 1961; Holland & Astin, 1962; Holland & Nichols, 1964; Holland & Richards, 1965; Nichols & Holland, 1963).

The most important of these findings is the low relationship between nonclassroom achievements and measures of academic potential and performance. These correlations are based, of course, on combining students at the various colleges into a single group. Although it is unlikely, this low relationship might be an artifact of combining students in different colleges. To check this possibility, the correlations between academic predictors and all criteria for male sophomores at individual colleges were computed. This analysis was restricted to the 14 colleges having 25 or more students with complete data.

The results indicate that there is indeed considerable variation among colleges in the relationship between individual predictors and individual criteria. However, the median correlations in every case are very close to the corresponding correlations which were calculated using all students combined. Moreover, the differences among colleges apparently are more random than consistent and meaningful. These results indicate, therefore, that combining students from different colleges has not distorted the relationships. Correlations at individual colleges for the other sam-

ples and other variables in this study supported this interpretation.

As our next step, we computed multiple correlations by selecting the most efficient predictors of each criterion from the 11 predictor variables. A step-wise multiple-regression computer program was used which, at each step, adds the variable which most improves prediction. This computer program computes an *F* test after each step to test the significance of the reduction of residual variance caused by the addition of the variable in that step. For the final multiple-regression equation, the computer retains only those variables producing a significant reduction in residual variance.

However, many variables which produced a statistically significant reduction in residual variance had no practical effect on the size of the multiple correlation. Accordingly, rather than using a statistical test, only those variables were retained which increased the multiple correlation by at least .01. In every case, the number retained using this criterion is smaller than the number retained using a statistical test of significance as the criterion.

Eight of the criterion variables—college grades, leadership, art, science, music, writing, speech and drama, and recognition for academic accomplishment—were designed specifically to assess at the college level the same characteristics the predictors measure at the high school level. The beta weights and multiple correlations for these criteria for freshmen are summarized in Table 2 and those for sophomores are summarized in Table 3.

The most notable finding in Tables 2 and 3 is the great importance of specific content in predicting achievement. For the nonacademic accom-

plishment scales, the best predictor of accomplishment in college is similar accomplishment in high school, and in the majority of cases similar high school accomplishment is the only variable contributing to the prediction of college accomplishment. Moreover, in every remaining case, the prediction of nonacademic accomplishment is improved only slightly by adding variables to the corresponding high school achievement scales—an improvement likely to disappear on cross-validation. These findings are consistent, of course, with a substantial literature which reveals that past performance predicts future performance.

For the two measures of academic accomplishment, the most consistently high predictor is high school grades, and, in general, some weighted combination of high school grades and ACT test scores is a better predictor than high school grades alone. This finding, too, is consistent with a large number of previous investigations of the prediction of academic performance.

The information in Tables 2 and 3 also confirms earlier findings that academic potential and success have little relationship to effective non-academic performance (Astin, 1962; Getzels & Jackson, 1962; Gough et al., 1963; Holland & Nichols, 1964;

TABLE 2

MULTIPLE CORRELATIONS FOR COLLEGE FRESHMEN FOR CRITERIA OF ACHIEVEMENT HIGHLY COMPARABLE TO THE HIGH SCHOOL ACHIEVEMENT SCALES

Criterion	Men			Women		
	Predictors	Beta	r	Predictors	Beta	r
College grades	ACT social studies	.2406	.29	High school grades	.2874	.37
	High school grades	.1316	.32	ACT social studies	.2580	.44
	Art achievement (HS)	.0838	.33			
Leadership achievement (Col.)	Leadership achievement (HS)	.2331	.29	Leadership achievement (HS)	.1934	.25
	Drama achievement (HS)	.1538	.32	Drama achievement (HS)	.1798	.30
Artistic achievement (Col.)	Art achievement (HS)	.3894	.41	Art achievement (HS)	—	.49
	Literary achievement (HS)	.0984	.42			
Scientific achievement (Col.)	Science achievement (HS)	—	.31	Science achievement (HS)	—	.22
Musical achievement (Col.)	Music achievement (HS)	.4157	.41	Music achievement (HS)	.3310	.35
	ACT natural science	-.0950	.42	Drama achievement (HS)	.0905	.36
Writing achievement (Col.)	Literary achievement (HS)	—	.43	Literary achievement (HS)	.3939	.44
				Drama achievement (HS)	.1439	.46
Speech & drama achievement (Col.)	Drama achievement (HS)	.3391	.34	Drama achievement (HS)	—	.44
	ACT natural science	-.0861	.35			
Recognition for academic accomplishment (Col.)	High school grades	.1841	.23	High school grades	.1079	.19
	Science achievement (HS)	.1386	.27	Leadership achievement (HS)	.1211	.25
	Leadership achievement (HS)	.1028	.29	ACT mathematics	.1376	.27
				Drama achievement (HS)	.0913	.29
				Science achievement (HS)	.0865	.30

Note.—In this and the following three tables, only variables increasing the multiple correlation by at least .01 are retained. The correlation shown beside each variable is the multiple correlation with the designated criterion of that variable plus those listed above it. Abbreviations in parentheses are as follows: Col. = College, HS = High School. For men, $N = 503$; for women, $N = 592$.

TABLE 3

MULTIPLE CORRELATIONS FOR COLLEGE SOPHOMORES FOR CRITERIA OF ACHIEVEMENT
HIGHLY COMPARABLE TO THE HIGH SCHOOL ACHIEVEMENT SCALES

Criterion	Men			Women		
	Predictors	Beta	r	Predictors	Beta	r
College grades	High school grades ACT social studies	.3168 .2091	.39 .44	High school grades ACT English ACT social studies	.3162 .1625 .1535	.43 .48 .50
Leadership achievement (Col.)	Leadership achievement (HS) Drama achievement (HS) Literary achievement (HS)	.1857 .1253 .1036	.28 .31 .33	Leadership achievement (HS) ACT social studies Literary achievement (HS)	.2984 .1283 .1092	.35 .38 .39
Artistic achievement (Col.)	Art achievement (HS)	—	.44	Art achievement (HS)	—	.51
Scientific achievement (Col.)	Science achievement (HS)	—	.40	Science achievement (HS)	—	.24
Musical achievement (Col.)	Music achievement (HS)	—	.49	Music achievement (HS)	—	.39
Writing achievement (Col.)	Literary achievement (HS)	—	.45	Literary achievement (HS)	—	.46
Speech & drama achievement (Col.)	Drama achievement (HS) Literary achievement (HS)	.2892 .0928	.33 .34	Drama achievement (HS) ACT mathematics	.3900 -.0900	.39 .40
Recognition for academic accomplishment (Col.)	High school grades ACT English Literary achievement (HS) ACT mathematics	.2113 .1276 .1466 .1470	.34 .39 .42 .43	High school grades ACT English ACT natural science	.2619 .1630 .1541	.37 .43 .45

Note.—Abbreviations in parentheses are as follows: Col. = College, HS = High School. For men, $N = 1373$; for women, $N = 1419$.

Hoyt, 1966; MacKinnon, 1960; Price et al., 1964; Thorndike & Hagen, 1959; Torrance, 1963). In these tables, academic predictors relate to academic criteria and nonclassroom predictors relate to nonclassroom criteria. Thus there is both convergent and discriminant validity. This is especially important in the case of the Recognition for Academic Accomplishment Scale. This scale is a self-report of achievements comparable to the nonclassroom achievement scales. Furthermore, the items for this scale were mixed with items from the nonclassroom achievement scales in the same section of the follow-up questionnaire. Unlike the nonclassroom achievement scales, however, this scale should be correlated with

academic predictors. Because this scale was correlated with academic predictors and the nonclassroom achievement scales were not, the results make it less plausible that response bias, dissimulation, or similar occurrences invalidate student responses to these scales. In other words, the results imply that the average student gave a frank account of his accomplishments in high school and in college.

The remaining six criterion scales make our assessment of student accomplishment more comprehensive; but they were not planned to measure achievement in the same areas measured by the high school achievement scales. It was expected, then, that the multiple correlations between the

criteria and the predictors would be lower than the correlations for the criteria that are highly comparable to the high school achievement scales. The multiple correlations for these criteria are summarized for freshmen in Table 4 and for sophomores in Table 5.

The multiple correlations in Tables 4 and 5 are much lower than the multiple correlations in Tables 2 and 3. In Tables 4 and 5, there is some tendency for those scales that are most similar to the high school achievement scales to be most predictable, and for the most similar high school scale to be the best predictor of the score on the similar college achievement scale. For example, high school Leadership Achievement is the best predictor of college Social

Participation, and high school Literary Achievement is the best predictor of college Humanistic-Cultural Achievement. For the most part, the correlations in Tables 4 and 5 support the conclusion that academic predictors contribute little to the prediction of nonclassroom accomplishment.

Again, probably the most striking thing suggested by Tables 4 and 5 is the importance of specific content. For the college criteria having no corresponding high school predictors, the variables selected for predicting the various criteria, and their beta weights, are not highly comparable for freshmen and sophomores. One would expect, therefore, the already low multiple correlations to drop on cross-validation. Consequently, a better approach to predicting these var-

TABLE 4
MULTIPLE CORRELATIONS FOR COLLEGE FRESHMEN FOR CRITERIA OF ACHIEVEMENT NOT HIGHLY COMPARABLE TO THE HIGH SCHOOL ACHIEVEMENT SCALES

Criterion	Men			Women		
	Predictors	Beta	r	Predictors	Beta	r
Social participation (Col.)	Leadership achievement (HS)	.2325	.28	Leadership achievement (HS)	.2482	.33
	Literary achievement (HS)	.1365	.31	Literary achievement (HS)	.1984	.38
	ACT social studies	.1717	.32	Science achievement (HS)	.1015	.40
	ACT English	-.1371	.34			
Social service achievement (Col.)	Drama achievement (HS)	.1048	.13	Leadership achievement (HS)	.1637	.21
	Music achievement (HS)	.1048	.17	Drama achievement (HS)	.1492	.25
Business achievement (Col.)	ACT English	-.1377	.14	ACT English	-.0800	.12
	Music achievement (HS)	-.0759	.16	Drama achievement (HS)	.0812	.14
				ACT natural science	-.0800	.16
Humanistic-cultural achievement (Col.)	Literary achievement (HS)	.2888	.33	Literary achievement (HS)	.2730	.33
	Drama achievement (HS)	.0958	.34	Art achievement (HS)	.1198	.35
				Drama achievement (HS)	.1107	.37
Religious service (Col.)	Music achievement (HS)	.1434	.14	Drama achievement (HS)	.1221	.13
	ACT English	-.1143	.18	ACT English	-.0620	.14
				Music achievement (HS)	.0581	.15
Social science achievement (Col.)	Literary achievement (HS)	.0915	.10	Literary achievement (HS)	.1470	.20
	ACT mathematics	-.1366	.15	Drama achievement (HS)	.1138	.23
	ACT social studies	.1794	.18	Art achievement (HS)	.0922	.25
	ACT natural science	-.1059	.20			

Note.—Abbreviations in parentheses are as follows: Col. = College, HS = High School. For men, $N = 503$; for women, $N = 592$.

TABLE 5

MULTIPLE CORRELATIONS FOR COLLEGE SOPHOMORES FOR CRITERIA OF ACHIEVEMENT NOT HIGHLY COMPARABLE TO THE HIGH SCHOOL ACHIEVEMENT SCALES

Criterion	Men			Women		
	Predictors	Beta	r	Predictors	Beta	r
Social participation (Col.)	Leadership achievement (HS)	.2340	.32	Leadership achievement (HS)	.2423	.30
	Literary achievement (HS)	.1404	.37	Literary achievement (HS)	.2008	.35
	High school grades	-.1597	.39	High school grades	-.0899	.36
	Drama achievement (HS)	.1371	.40			
	ACT social studies	.1057	.42			
Social service achievement (Col.)	Leadership achievement (HS)	.1825	.25	Leadership achievement (HS)	.1342	.21
	Drama achievement (HS)	.1338	.29	Drama achievement (HS)	.1260	.25
	High school grades	-.1277	.32	Science achievement (HS)	.0938	.26
	Literary achievement (HS)	.0911	.33			
Business achievement (Col.)	Science achievement (HS)	.1373	.17	Drama achievement (HS)	.0985	.13
	Drama achievement (HS)	.0834	.20	Leadership achievement (HS)	.0820	.15
	High school grades	-.1020	.22	ACT social studies	-.0617	.16
	Leadership achievement (HS)	.0929	.24			
Humanistic-cultural achievement (Col.)	Literary achievement (HS)	.2974	.35	Literary achievement (HS)	.1899	.29
	ACT social studies	.2073	.38	ACT social studies	.1453	.32
	ACT mathematics	-.1285	.40	Drama achievement (HS)	.1257	.35
	Art achievement (HS)	.1072	.41	Art achievement (HS)	.1086	.36
Religious service (Col.)	Drama achievement (HS)	.1135	.15	Music achievement (HS)	.1134	.13
	Leadership achievement (HS)	.0804	.16	ACT social studies	-.0859	.15
	ACT social studies	-.0609	.18	Drama achievement (HS)	.0611	.16
Social science achievement (Col.)	Literary achievement (HS)	.1330	.17	Literary achievement (HS)	.1729	.19
	Leadership achievement (HS)	.1414	.20	Art achievement (HS)	.1006	.21
	High school grades	-.1203	.23			

Note.—Abbreviations in parentheses are as follows: Col. = College, HS = High School. For men, $N = 1373$; for women, $N = 1419$.

ables would seem to be to construct a high school achievement scale corresponding closely to the college achievement scale. When predictors are available which are expected to have substantial validity on rational grounds and on the basis of previous research, as was the case with the highly comparable high school and college achievement scales in this study, they may not necessarily be improved (on cross-validation) by adding variables selected from a large number of predictors to maximize the multiple correlation. Indeed, because the multiple correlation may

weight the single, dependable predictor inappropriately in the process of combining it with other variables, the validity of the weighted combination may actually be lower than the validity of the single variable alone in a new sample (Holland & Nichols, 1964).

DISCUSSION

The present study demonstrates that it is possible to predict nonacademic accomplishment with moderate success, and it extends the similar findings of earlier research on students with high aptitude by showing

that this is true for students with a broad range of academic potential. To illustrate, the median correlation between student nonacademic accomplishment in high school and in college in the same area of endeavor is about .39; the median correlation between ACT scores and college grades is about .29; and the median correlation between grades in high school and in college is about .38. These values are not strictly comparable, of course, for at least two reasons: many students in the original sample left college because of low grades; and we did not correlate individual ACT tests with grades in specific courses. Nevertheless, the results suggest that the predictive validities of the high school accomplishment scales⁴ are about as high for comparable criteria as the predictive validities of the ACT tests.

This study, therefore, is the culmination of our research to establish that some nonacademic accomplishments are independent of academic potential and accomplishment (Holland & Richards, 1965, in press), that nonacademic accomplishment can be assessed with moderate reliability (American College Testing Program, 1965; Richards et al., in press), and that nonacademic potential can be predicted with moderate success (Holland & Nichols, 1964). The evidence also makes it unlikely that our results can be attributed to nonlinear relationships between academic and nonacademic accomplishment (Holland & Richards, 1965), to defective scaling of nonacademic accomplishments (Holland & Nichols, 1964; Holland & Richards, 1965), to a narrow range of student talent (Holland & Richards, 1965, in press), to a student's distortion of his nonaca-

demie accomplishment (Holland & Richards, in press; Richards et al., in press), or to the effects of some moderator variables (Holland & Richards, in press). These results also support many of the findings of investigators of creative and effective performance (Gough et al., 1963; MacKinnon, 1960; Price et al., 1964; Thorndike & Hagen, 1959; and others). The recent review by Hoyt (1966) provides still another important piece of evidence that classroom grades bear little or no relationship to measures of adult accomplishment.

Because our criteria of nonacademic accomplishment are only a sample of such accomplishment, possibly academic potential and accomplishment may have substantial positive correlations with *some* nonacademic accomplishments. The negligible relationships observed so far, however, make this possibility unlikely. While only an exhaustive examination of nonacademic accomplishments could negate this possibility, some relevant evidence is provided by the six new criteria of nonacademic accomplishment⁵ developed for this study. The negligible relationships between measures of academic potential and performance and these new criteria of nonacademic accomplishment reinforce earlier findings and lessen the possibility of finding some substantial positive correlations.

As always, the present research leaves a number of closely related questions unanswered. It is not yet known whether nonclassroom accomplishments in high school and college are good predictors of similar accomplishment in adult life. Little is known about the college experiences that

⁴For the following six scales: Science, Art, Music, Literary, Drama, and Leadership.

⁵These criteria are: Social Participation, Social Service Achievement, Business Achievement, Humanistic-Cultural Achievement, Religious Service, and Social Science Achievement.

facilitate and inhibit the expression of talent in college after a record of talented performance is made in high school. The apparent contradictions between the findings of Terman and Oden (1959) and the findings of more recent investigations, such as the present study, need to be resolved. Similarly, the relationship of such work as Thurstone's (1938) primary mental abilities and Vernon's (1950) hierarchy of abilities to nonacademic accomplishment requires explication. A theory of human accomplishment encompassing our notions of intelligence, aptitude, nonacademic accomplishment, and originality would help us find answers to these questions.

Some of the practical applications of the findings seem clear. Measures of academic potential are the chief methods used to determine admission to college (Committee on School and College Relations, 1964). So long as one is interested *only* in finding students who will do well in the classroom in college, this emphasis is appropriate. But the emphasis in colleges and universities on academic potential, because it concentrates on only one of several independent dimensions of talent, has led to neglect of other equally important talents. Certainly, in the interest of social and human values, one should also be interested in finding students who will do outstanding things outside the classroom and in later life.

We should, therefore, continue to develop and improve measures of many kinds of achievement and of originality. Further, we should consider such measures important in their own right, and not weak supplementary measures to remedy the slight defects of conventional aptitude and achievement tests. At the same time, we should not make the same mistake that the proponents of aptitude and

intelligence have made in the past; that is, to rely on only one kind of measure and to exclude others. The results support some of the items used to obtain information about non-classroom accomplishment in typical application blanks for admission to college, scholarships, and fellowships, but they also suggest the potential usefulness of a more reliable and valid record of each student's past achievement and involvement.

The implications of this study, however, extend beyond a need for a more systematic and comprehensive assessment of student accomplishment outside the classroom for purposes of admission or selection. At the very least, the findings imply a need to examine college grading practices, since college education should be largely a preparation for participation in important areas of human endeavor. Because college grades best predict graduate grades, current grading practices imply that a college education is mainly preparation for more education in graduate school. The criteria of nonacademic accomplishment, in combination with college grades, provide a brief set of socially relevant measures which could serve as more comprehensive criteria of college success. Using these scales as guides, similar scales can be developed to increase our ability to assess student attainment of the broader goals of a college education. Moreover, once the simple principles of constructing such scales are grasped, it should be easy to develop scales to satisfy a particular college's unique needs.

Further, the results imply a need for a broader, or different, definition of both the nature of human talent and the nature of higher education. There are many kinds of human accomplishment, and each kind is likely

to benefit from some type of higher education, although not necessarily a highly academic type. In other words, our results imply a need for a wide variety of colleges, many, if not most of them, relatively unselective except on dimensions clearly relevant to their particular emphasis. Measures of academic and nonacademic accomplishment would then be used in helping students find an appropriate college, rather than being used in selecting students for a single college.

As one critic of education said, a society (or a system of higher education) is "in a desperate way when its music makes little difference [Goodman, 1966]." Despite contrary protestations, most institutions of higher education rely heavily on academic aptitude and grades in selecting and evaluating students. Music and other important human accomplishments make little difference.

REFERENCES

- ABE, C., HOLLAND, J. L., LUTZ, S. W., & RICHARDS, J. M., JR. *A description of American college freshmen*. ACT Research Report No. 1. Iowa City: American College Testing Program, 1965.
- AMERICAN COLLEGE TESTING PROGRAM. *ACT technical report*. Iowa City: Author, 1965.
- ASTIN, A. W. *A note on the relationship between creative and academic achievement*. Evanston, Ill.: National Merit Scholarship Corp., 1962.
- BARRON, F. *Creativity and psychological health*. Princeton, N. J.: Van Nostrand, 1963.
- BUEL, W. D. Biographical data and the identification of creative research personnel. *Journal of Applied Psychology*, 1965, **45**, 318-321.
- CHAMBERS, J. A. Relating personality and biographical factors to scientific creativity. *Psychological Monographs*, 1964, **78**(7, Whole No. 584).
- CICIRELLI, V. G. Form of the relationship between creativity, I. Q., and academic achievement. *Journal of Educational Psychology*, 1965, **56**, 303-308.
- COMMITTEE ON SCHOOL AND COLLEGE RELATIONS. *Admission to American colleges*. New York: Educational Records Bureau, 1964.
- DAVIDSEN, O. M. Reliability of self-reported high school grades. Unpublished research report, American College Testing Program, 1963.
- EELLS, K. *How will the ACT tests serve Illinois needs?—A research report*. Committee on Freshman Testing, Illinois Joint Council on Higher Education, 1962.
- FLESCHER, I. Anxiety and achievement of intellectually gifted and creatively gifted children. *Journal of Psychology*, 1963, **56**, 251-268.
- GETZELS, J. W., & JACKSON, P. W. *Creativity and intelligence: Explorations with gifted students*. New York: Wiley, 1962.
- GOODMAN, P. *Compulsory miseducation and the community of scholars*. New York: Vintage, 1966.
- GOUGH, H. G., HALL, R. E., & HARRIS, W. B. Admissions procedures as forecasters of performance in medical education. *Journal of Medical Education*, 1963, **38**, 983-998.
- GUILFORD, J. P. Some new looks at the nature of creative processes. In N. W. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rhinehart, and Winston, 1964. Pp. 161-176.
- HOLLAND, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *Journal of Educational Psychology*, 1959, **50**, 135-142.
- HOLLAND, J. L. The prediction of college grades from personality and aptitude variables. *Journal of Educational Psychology*, 1960, **51**, 245-254.
- HOLLAND, J. L. Creative and academic performance among talented adolescents. *Journal of Educational Psychology*, 1961, **52**, 136-147.
- HOLLAND, J. L., & ASTIN, A. W. The prediction of the academic, artistic, scientific, and social achievement of undergraduates of superior scholastic aptitude. *Journal of Educational Psychology*, 1962, **53**, 132-143.
- HOLLAND, J. L., & NICHOLS, R. C. Prediction of academic and extracurricular achievement in college. *Journal of Educational Psychology*, 1964, **55**, 55-65.
- HOLLAND, J. L., & RICHARDS, J. M., JR. Academic and nonacademic accomplishment: Correlated or uncorrelated? *Journal of Educational Psychology*, 1965, **56**, 165-174.
- HOLLAND, J. L., & RICHARDS, J. M., JR. Aca-

- demic and non-academic achievement in a representative sample of students taking the American College Tests. *College and University*, in press.
- HOYT, D. P. Three years of research in the American College Testing Program. Paper presented at the meeting of the Iowa Psychological Association, Des Moines, May, 1964.
- HOYT, D. P. College grades and adult accomplishment: A review of research. *The Educational Record*, 1966, Winter, 70-75.
- LOCKE, E. A. The development of criteria of student achievement. *Educational and Psychological Measurement*, 1963, 23, 299-307.
- MACKINNON, D. W. What do we mean by talent and how do we test for it? In, *The search for talent*. New York: College Entrance Examination Board, 1960. Pp. 20-29.
- MANN, M. J. Relationships among certain variables associated with college and post-college success. Unpublished doctoral dissertation, University of Wisconsin, 1958.
- MUNDAY, L. *Comparative predictive validities of the American College Tests and two other scholastic aptitude tests*. ACT Research Report No. 6. Iowa City: American College Testing Program, 1965.
- NICHOLS, R. C., & HOLLAND, J. L. Prediction of the first year college performance of high aptitude students. *Psychological Monographs*, 1963, 77(7, Whole No. 570).
- PRICE, P. B., TAYLOR, C. W., RICHARDS, J. M., JR., & JACOBSEN, T. L. Measurement of physician performance. *Journal of Medical Education*, 1964, 39, 203-211.
- RICHARDS, J. M., JR., HOLLAND, J. L., & LUTZ, S. W. The assessment of student accomplishment in college. *The Journal of College Student Personnel*, in press.
- SKAGER, R. W., SCHULTZ, C. B., & KLEIN, S. P. Quality and quantity of accomplishments as measures of creativity. *Journal of Educational Psychology*, 1965, 56, 31-39.
- SPRECHER, T. B. A study of engineers' criteria for creativity. *Journal of Applied Psychology*, 1959, 43, 141-148.
- TAYLOR, C. W., SMITH, W. R., & GHISELIN, B. The creative and other contributions of one sample of research scientists. In C. W. Taylor & F. Barron (Eds.), *Scientific creativity: Its recognition and development*. New York: Wiley, 1963. Pp. 228-250.
- TERMAN, L. M., & ODEN, M. H. *Genetic studies of genius. V. The gifted group at midlife*. Stanford: Stanford University Press, 1959.
- THORNDIKE, R. L., & HAGEN, E. *10,000 careers*. New York: Wiley, 1959.
- THURSTONE, L. L. Primary mental abilities. *Psychometric Monographs*, 1938, No. 1.
- TORRANCE, E. P. *Education and the creative potential*. Minneapolis: University of Minnesota Press, 1963.
- VERNON, P. E. *The structure of human abilities*. London: Methuen, 1950.
- WALLACH, M. A., & KOGAN, N. *Modes of thinking in young children: A study of the creativity-intelligence distinction*. New York: Holt, Rhinehart, and Winston, 1965.

(Received November 21, 1966)

LEARNING VERBAL AND SYMBOLIC STATEMENTS OF MATHEMATICAL RULES¹

JOSEPH M. SCANDURA

University of Pennsylvania

24 college Ss learned 4 rule statements and were then tested on 2 problems to see if they could apply the rules involved. A 2×2 factorial design, with repeated measures, was used. 1 factor was the form in which the rules were stated, either succinctly worded English or initially unfamiliar mathematical symbolism. The other factor involved the presence or absence of symbol pretraining on the meaning of the constituent symbols. The results were: (a) symbolic statements were applied successfully if and only if the Ss were taught the symbol meanings (and the underlying grammar); (b) symbolic statements were learned more rapidly with or without symbol pretraining; (c) English statements were applied equally as well as those symbolic statements which followed symbol pretraining; (d) rate of learning the English statements was unaffected by symbol pretraining; (e) success on 1 application problem implied success on the other.

In recent years, there has been increasing concern with the proposition that the effectiveness of a given mode of presentation depends on learner ability. This supposition was originally interpreted to mean that aptitude profiles could be used to predict which mode of instruction would be most effective in promoting learner achievement. Thus, people who are high in spatial but low in verbal and symbolic ability, for example, presumably would do best when concepts are presented in spatial form. Similar interactions would be expected with different profiles.

Gagné (1960) was one of the first to propose this sort of interaction hypothesis with respect to mathematics. Although there have been some studies since then which tend to support this point of view (e.g., Guilford, Hoepfner, & Peterson, 1965; Osburn & Milton, 1963), most of the studies indicate that aptitude profiles are not

nearly as strong predictors of learning as are immediately requisite abilities. In dealing with hierarchially ordered tasks, for example, Gagné and his collaborators (Gagné, 1962; Gagné, Mayor, Garstens, & Paradise, 1962; Gagné & Paradise, 1961) found that the relationship between successful criterion performance and ability, as measured by aptitude tests, became progressively weaker as learning moved towards the hierarchical apex. Conversely, the relationship between criterion performance and acquired subordinate knowledge, as determined during the learning sequence, became progressively stronger. Largely as a result of these studies, the general validity of the task-analysis procedure, introduced in education by Gagné and his co-workers, has become so well accepted that the basic procedure has been extended by others—for example, to deal with *process* as well as *content* objectives (Kersh, 1967).

Unfortunately, however, there has been no attempt to vary the form in which new material (whether relating to content or process) is presented.

¹ This research was supported in part by the United States Office of Education. The author would like to thank John Davis for his assistance with several phases of the experiment.

The important question remains as to whether aptitude measures, while perhaps not as critical to learning as immediately requisite abilities, may still be the best predictors of which "form of content" should be used.

In a recent study, Scandura (1966a) found that the repeated reintroduction of passages of technical material did not diminish the initial advantage gained by those subjects (*Ss*) who had previously been required to demonstrate a high proficiency with the terminology used. This result obtained even though all of the *Ss* were exposed, in varying but parallel degrees, to definitions of and practice with the prerequisite terms before the technical material was initially introduced. In discussing the results of their study on rule generality, Scandura, Woodward, and Lee (1967) also suggested that the ability to interpret a given rule statement may depend critically on the terms used to compose the statement. It would appear that specific interpretive abilities may be more fundamental than any general aptitude measure, no matter in what form information is presented. More particularly, the interpretation of a statement may depend on certain requisite abilities which can be logically determined directly from the to-be-presented material itself in a manner analogous to that used in task analysis (Gagné, 1962; Kersh, 1967)—that is, by asking the question, "What does *S* need to be able to do in order to interpret . . . ?"

It was with this general orientation in mind that the present study was designed to help clarify the role symbolism plays in learning mathematical rules. (See Scandura, 1966c, 1967; Scandura et al., 1967, for a precise definition of a rule.) Both the symbols actually used to construct rule statements and the ability of an *S* to in-

terpret symbols were varied. It seems almost axiomatic that the ability to interpret a rule statement depends critically on the ability to interpret the symbols of which it is composed (Scandura, 1966a), be they mathematical symbols or elements of the native language (e.g., English). Whether symbol meanings are a sufficient condition for rule interpretability, in the deterministic sense, however, is not at all clear. Furthermore, in mathematics learning the use of mathematical symbolism is frequently, if not always, preferred to ordinary English. The reason why, however, is never made explicit.

The purpose of this study was to determine whether: (a) rules are more easily memorized when stated in mathematical symbolism or when stated verbally (i.e., in English), and (b) the ability to use constituent symbols correctly, assuming mastery of the underlying grammar, is a necessary and/or sufficient condition for applying a learned (i.e., memorized) rule statement.

There are, of course, innumerable ways in which a given rule may be stated and it would be a tedious task indeed to make all possible experimental comparisons. Unless general relationships can be found without making all (or even many) such comparisons, research on the question of whether rules are more easily memorized in symbolic or verbal form would be as fruitless and as hopelessly complex as traditional methods research in which two or more ill-defined instructional procedures are compared.

Hopefully, the present study does not fit that classification. An independent variable, called description level, has been defined (Scandura, 1966c) which may be directly related to the rate of learning statements of a given rule. *Statement A is said to be*

of a higher description level than *B* with respect to some reference symbolism (usually the native language) if the translation of *A* into the reference symbolism requires all those reference symbols needed to translate *B* plus some additional symbols of its own. Statements generally can be only partially ordered according to this definition. That is, some statements may not be directly comparable as to description level. Nonetheless, if a positive relationship is found between description level and rate of learning, as hypothesized in this study, it should be possible, by interpolation and/or extrapolation, to make confident predictions as to the relative rates of learning new rule statements which take on unstudied values of description level. In the present study, two values of the description level variable were compared, statements in the native language (i.e., English) and statements of identical rules composed of mathematical symbols.

Data regarding the second purpose of the study were obtained after all of the rule statements were well learned. In this regard, a secondary result of research by Scandura and his collaborators (Scandura, 1966b, 1966c, 1967a; Scandura et al., 1967) suggests that, under certain conditions, any one instance of a rule would provide an adequate test of interpretability. It was found that if the appropriate response could be given to one stimulus within the scope of a rule, then the appropriate response to any other within-scope stimulus could also be given. In short, no additional information was gained by using more than one test instance. Even so, two test instances were used in the present study, thereby making it possible to gain further information on the implied consistency hypothesis.

METHOD

Design and Subjects

A 2×2 factorial design, with repeated measures, was used so that each of the 24 *Ss* effectively served as his own control. One factor was the form in which the rules were stated, either succinctly worded English or mathematical symbolism. The other factor involved the presence or absence of symbol pretraining on the meaning of the constituent mathematical symbols. The four rules (eight statements) used were counter-balanced over the four treatment-combinations so that each of the 24 (4!) ways of assigning 4 rules to 4 treatments was used once—one permutation per *S*. All other factors were randomized, including presentation order.

The 22 female and 2 male *Ss* were Florida State University elementary education majors taking a methods course in mathematics. Participation was not required but was encouraged.

Materials and Procedures

The *Ss* reported for the experiment one at a time. The *S* was told,

This experiment is designed to determine how people learn rules for solving mathematical problems. I am going to present some rules for you to learn. After you learn these rules, I will give you some mathematical problems to see if you can apply these rules.

Before learning the rules, *S* was given some common pretraining to be sure that he understood the way certain terms were to be used. In particular, he was told: (a) to perform operations within the innermost parentheses first before performing additional operations, (b) what the consecutive integers from 1 through *k* were, and (c) that

$$\begin{vmatrix} x_1 \\ x_2 \end{vmatrix}$$

was a vector with x_1 in the first position (row) and x_2 in the second position, (d) that $ax^m y^n$ is an algebraic expression, and (e) that the greatest integer in a number is the largest integer less than or equal to that number. In each case, *S* was presented with one illustration and was required to complete, in a correct manner, four practice exercises based directly on the information given. None of this information was deemed

TABLE 1
RULE STATEMENTS

Name of rule	Symbolic form	Verbal form
Greek	$\sum_{s=1}^i z \left(\sum_{x=1}^r x + \sum_{y=1}^s y \right)$	1. Add the consecutive integers from 1 thru r. 2. Add the consecutive integers from 1 thru s. 3. Add the two sums obtained in Steps 1 and 2. 4. Add the consecutive integers from 1 thru t. 5. Multiply the sum resulting from Step 3 by the sum resulting from Step 4.
Integer	$[[x] \div [y]]$	1. Take the greatest integer in x. 2. Take the greatest integer in y. 3. Divide the result of Step 1 by the result of Step 2. 4. Take the greatest integer in the quotient in Step 3.
Vector	$\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \odot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) + \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \odot \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right)$	1. Multiply the first number in $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ by the first number in $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. 2. Multiply the second number in $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ by the second number in $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. 3. Add the two products obtained in Steps 1 and 2. 4. Multiply the first number in $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ by the first number in $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$. 5. Multiply the second number in $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ by the second number in $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$. 6. Add the two products obtained in Steps 4 and 5. 7. Add the two products obtained in Steps 3 and 6.
Partial	$D_y[D_x(ax^m y^n)]$	1. Multiply the given expression $ax^m y^n$ by the exponent of x. 2. Reduce the exponent of x in the algebraic expression resulting from Step 1 by 1. 3. Multiply the algebraic expression resulting from Step 2 by the exponent of y. 4. Reduce the exponent of y in the algebraic expression resulting from Step 3 by 1.

sufficient for understanding any of the symbolic rule statements.

Next, S was given specific information about the mathematical symbols used to compose two of the four rule statements (see Table 1) later introduced. He was re-

quired to learn this information before going on. In each case, the presumably unfamiliar symbols used to compose the rules were defined, an example was given, and four tasks, similar to the example, were presented for S to practice what he had been

told. The definition corresponding to the Greek rule was stated:

To find the number equal to

$$\sum_{r=1}^k r$$

add the consecutive integers from 1 thru k .

The illustration was

$$\sum_{t=1}^6 t = 1 + 2 + 3 + 4 + 5 + 6 = 21$$

and the practice problems were identical in form. The other definitions were stated:

Integer—the brackets about the numeral A , $[A]$, mean take the largest integer less than or equal to A , e.g., $[3.57] = 3$;

Vector—if

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

are vectors, then to get the number equal to

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \odot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix};$$

(a) multiply the first number in

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

by the first number in

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

(b) multiply the second number in

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

by the second number in

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

(c) add the two products obtained in Steps a and b, e.g.,

$$\begin{bmatrix} 6 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 6 \times 3 + 5 \times 4 = 18 + 20 = 38;$$

Partial—to find $D_s(ax^m y^n)$

(a) multiply $ax^m y^n$ by m

(b) reduce the exponent of x in the algebraic expression resulting from Step a by 1, for example, to find the algebraic expression equal to $D_s(5x^4 y^6)$; Step a gives $20x^4 y^6$, Step b gives $20x^3 y^6 = D_s(5x^4 y^6)$.

Just before the rules were presented, S was told,

The name of each rule, an introductory statement, and the corresponding rule have been typed on a card. The name and introductory statement have been typed in black and the rule you are to learn in red. Your job will be to memorize that part typed in red so that you can write the rule correctly whenever you see its name.

You will have a given period of time to study each rule statement that I show you. After you have studied each rule, I will present each name, in turn. Your job will be to learn all four rules so that you can write the rule statement corresponding to each of the four names shown. You will have all the time you need but please work rapidly. If you haven't learned the rule, write as much of it as you can and we will then go on.

To be sure that S understood the procedure, the experimenter presented four distinct sample rules in the manner indicated. Then, S was told that he would have to write each rule several times without a mistake and that more time would be allowed on successive trials. The S was also reminded that he would be required to apply what he had learned.

The order of presentation was randomized on each trial and S was told whether he was right or wrong after each attempt to reproduce a rule statement. The test phase of each trial took place after all four rules were presented. On Trial 1, 5 seconds were allowed for the study of each rule; on Trial 2, 10 seconds; on Trial 3, 15 seconds; and so on. To minimize the time required for writing while insuring that each rule statement was well learned, those rules which S successfully reproduced twice in a row were not reintroduced until all of the others had been learned to the same criterion. As a final review, S went through all four rules until he made no mistakes. In each case, S did this on the first (review) trial.

After the learning criterion had been met, S was required to apply each rule to two stimulus instances. He was told,

Each problem will appear together with the name of one of the rules you have learned. You are to simply apply this rule to the problem shown. That is, write the correct answer.... You may refer to any of the information you have already seen any time you need it.

In particular, the four rules were in full view of S throughout the *interpretation-test*

period. The order of testing on the eight problems was random except that each rule was tested once before any other was tested a second time. All of the testing was self-paced. Finally, *S* was asked how he remembered each rule, whether he found any shortcuts, and to make any other comments he cared to.

All of the materials were reproduced by ditto and *S* wrote all of his answers—rules during the learning and solutions during the interpretation test—in the spaces provided.

In summary, separate measures of learning rate and interpretability were obtained. Learning rate was determined by presenting each rule statement for study and testing on each trial to see if *S* could correctly reproduce them in written form. Number of trials to a criterion of two perfect reproductions in a row was the dependent measure. Interpretability was measured by having *S* apply each of the rules taught to two stimulus instances. The interpretability testing took place immediately after all of the statements had been learned.

RESULTS

The major results were straightforward and are summarized in Table 2.

Those rules, which were stated in symbolic form, were applied successfully *if and only if* *S* had been taught how to apply the constituent symbols—assuming mastery of the underlying grammar (i.e., use of parentheses). There were only four of 24 (one per *S*) consistent exceptions, where neither problem was solved, to the sufficiency part of this generalization and none as regards necessity. The four exceptions were spread out over three of the four rules. Two other *S*s each failed to give the correct response to one of the two applications problems.²

²To clarify the underlying logic of this result, let *m* and $\sim m$ represent knowledge and lack of knowledge of the *symbol* meanings, and *a* and $\sim a$, ability and lack of ability to *apply* the corresponding rule, respectively. The (near deterministic) sufficiency result may be represented $m \rightarrow a$ (read *m* implies *a*) while the other demonstrated result may be represented $\sim m \rightarrow$

TABLE 2
LEARNING RATE AND APPLICABILITY OF
RULE STATEMENTS

	Number of trials to learning criterion		Number of correct applications	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Symbol form-symbols meaningful	4.83	1.63	1.58	.76
Symbol form-symbols nonmeaningful	4.92	1.10	0.00	.00
Verbal form-symbols meaningful	6.75	1.23	1.83	.57
Verbal form-symbols nonmeaningful	7.12	2.24	1.75	.66

The symbolic rule statements were learned more rapidly than the verbal statements [$F = 64.21$, $df = 1/23$, $p < .001$].³ More important, the symbolic statements were easier to learn whether or not the symbols were meaningful (F for the interaction was < 1). There were a total of only eight of 48 (two per *S*) exceptions to the latter two generalizations. In four of the eight situations, the symbolic and verbal statements were learned at the same rate.

Although they took longer to learn, the verbal statements were applied equally as well as (in fact, slightly better than) those symbolic statements in which use of the constituent

$\sim a$ which is equivalent to the contrapositive $a \rightarrow m$ (which is the necessary condition).

Also notice that, while each *S* served as his own control, the idiographic comparisons above and throughout the results section were necessarily averaged over different rules.

³This result was consistent and reliable for all but the *Greek* rule [$t(11) > 3.0$, $p < .01$]; in that case the difference was in the expected direction but was not significant ($t < 1$). The reason for this exception was not immediately clear but it may have been due to the relative complexity of the primary symbol,

$$\sum_{i=1}^r$$

which involved both sub- and superscripts.

symbols had previously been mastered. Furthermore, these verbal statements were learned at approximately the same rate and the underlying rules were applied equally well whether or not training in the use of the corresponding mathematical symbols was given. Only two of the 24 Ss were unable to apply the verbally stated rules when the symbols were meaningful and only three were unable to do so when they were not.

Strong support was again found for the consistency hypothesis. In only two cases out of 96 (four per S) was one of the two application problems solved and not the other.

DISCUSSION

At first glance, the effects of symbol pretraining do not appear to be particularly interesting or enlightening. Without an assigned referent, the constituent mathematical symbols were basically no different from nonsense syllables and could hardly be expected to be useful in applying a rule.

On the other hand, the near deterministic *sufficiency* data have important implications. Mastery of the constituent mathematical symbols—also assuming mastery of the underlying grammar (i.e., use of parentheses) and an appropriate experimental context—accounted for about 80% of the experimental outcomes on the interpretability tests. Much of the remaining 20% can probably be attributed to mistakes in computation, carelessness, and other momentary lapses.

What this means, of course, is that general mathematical aptitude and achievement measures can be expected to account for the ability to interpret rules, stated with mathematical symbols, only to the extent that they *covary* with the sort of interpretive prerequisites identified here. Ability

× Treatment interactions, where they exist and at least with respect to interpreting simple mathematical rules, can be attributed to very specific prerequisites. On the other hand, while viewing the ability-treatment question in this manner could have a profound effect on how the problem is attacked theoretically, it does not preclude the possibility that a demonstrated interaction, between some general aptitude or achievement measure and some mode of instruction, might have important practical implications. The very fact that most existing standardized tests get at a wide variety of specific skills and abilities makes it quite possible that the prerequisites for a given specific interpretive task will be sampled. Nonetheless, the results of this research suggest that better-defined tests may be constructed by identifying the nature of the information to be interpreted, determining and classifying the kinds of interpretive prerequisites required, and sampling these prerequisites across classifications to construct the predictive tests. I am not suggesting that this would be an easy task, but I am suggesting that a rationale-based approach to test construction may be a better way to deal with the ability-treatment question than an approach based purely on pragmatics.⁴

Why the symbolic rule statements were more easily learned than the verbal statements is a difficult question to answer definitively. While resulting in shorter statements, the

⁴ Although this is not the place for a lengthy discussion, similar arguments may be made regarding test construction in general. The research cited in the introduction suggests that it might be better to work backwards from the criterion task(s) in constructing tests rather than by following the more usual empirical procedure of constructing tests and later determining their validity.

mathematical symbols were not as familiar to the Ss as the corresponding verbal symbols. Still, with the materials used in this study it would appear that relative length was the more critical factor. In Miller's (1956) classic paper it was shown that man can increase his finite and relatively constant capacity for processing information by recoding information into a smaller number of "chunks." One likely manifestation of recoding is the mathematician's tendency to substitute more compact mathematical symbols for longer verbal phrases as their use becomes more frequent. According to this interpretation, learning the longer verbal statements probably required a substantially greater degree of recoding than did the others. The shorter symbolic statements, although composed of less familiar mathematical symbols, were effectively recoded to some degree already.

Providing S with the meanings of the mathematical symbols before he learned the corresponding verbal statements, however, did not reliably increase learning rate. This suggests that Ss tended to use their own preferred bases for recoding the verbal statements rather than the relatively unfamiliar bases (i.e., mathematical symbols) provided.⁵

Although the verbal rule statements took more time to learn than the symbolic statements, once learned, they were applied equally as well—in fact, slightly better. This result provided further support for an earlier finding by Scandura and Behr (1966). Apparently, the inherent redundancy of English, although resulting in relatively long rule statements, in no way

interfered with the application of the rules. It is possible, of course, that the application of particularly cumbersome rule statements, even when perfectly memorized, may require abilities beyond those had by many learners. Nonetheless, it may be well to remember that ordinary English, although less efficient, can be effectively used to teach precise mathematical ideas when appropriate mathematical symbolism is not immediately available to the learner.

While this study clearly does not deal with all that is relevant, these results do make explicit at least certain aspects of the role symbolism plays in learning mathematical rules. The use of mathematical symbols apparently makes learning (with "understanding") more efficient when the constituent symbols and underlying grammar have been previously mastered.

An extrapolation of the obtained results provides a rational and explicit basis for making one type of branching decision in instructional sequences—the sort of decision rule on which the effectiveness of computer-assisted instruction (CAI) may well rest. Given the objective of learning a particular rule and an expository mode of instruction, one might proceed as follows: (a) Test to see if S can make use of the constituent symbols, (b) if so, present the rule in the more efficient symbolic form, (c) if not, present the same statement in English. The circumstances under which this decision rule can be applied, of course, are extremely limiting. Nonetheless, it may be a step in the right direction. Heretofore, although student feedback has provided the basis for all adaptive instructional procedures, it has typically been unclear as to what sort of feedback to measure.

⁵ It may be just this sort of preference which resulted in a conceptual organizer having an unexpected negative effect on short-term memory (Scandura & Roughead, 1967).

The present results suggest that specific sorts of feedback are needed in order to make specific kinds of decisions. If this is true, it is quite possible (I feel quite likely) that general feedback measures, such as number of errors, average latency, etc., will play a less and less important role in CAI in the years ahead. Equally important, the obtained consistency results provide further reason to suspect that appropriate feedback may be determined in a more efficient manner than has been the case to date. Even highly efficient testing takes time, however, and, whereas testing is an integral part of the educational process, it must be used judiciously. There may be some optimal trade-off between the time devoted to testing and the time devoted to instruction. The role of basic research will be to help clarify these issues.

REFERENCES

- GAGNÉ, R. M. Implications of some doctrines of mathematics teaching for research in human learning. In *Research problems in mathematics education*. Washington: United States Office of Education (Cooperative Research Monograph, OE-12008), 1960.
- GAGNÉ, R. M. The acquisition of knowledge. *Psychological Review*, 1962, **69**, 355-365.
- GAGNÉ, R. M., MAYOR, J. R., GARSTENS, H. L., & PARADISE, N. E. Factors in acquiring knowledge of a mathematical task. *Psychological Monographs*, 1962, **76**(7, Whole No. 526).
- GAGNÉ, R. M., & PARADISE, N. E. Abilities and learning sets in knowledge acquisition. *Psychological Monographs*, 1961, **75** (14, Whole No. 518).
- GUILFORD, J. P., HOEFFNER, R., & PETERSON, H. Predicting achievement in 9th-grade mathematics from measures of intellectual-aptitude factors. *Educational & Psychological Measurement*, 1965, **25**, 659-682.
- KERSH, B. Y. Engineering instructional sequences for the mathematics classroom. In J. M. Scandura (Ed.), *Research in mathematics education*, 1967. Washington: National Council of Teachers of Mathematics.
- MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81-97.
- OSBURN, H. G., & MILTON, R. S. Prediction of proficiency in a modern and traditional course in beginning algebra. *Educational & Psychological Measurement*, 1963, **23**, 277-288.
- SCANDURA, J. M. Prior learning, presentation order, and prerequisite practice in problem solving. *Journal of Experimental Education*, 1966, **34**, 12-18. (a)
- SCANDURA, J. M. Teaching—technology or theory? *American Educational Research Journal*, 1966, **3**, 139-146. (b)
- SCANDURA, J. M. Precision in research on mathematics learning, the emerging field of psycho-mathematics. *Journal of Research in Science Teaching*, 1966, **4**, 253-274. (c)
- SCANDURA, J. M. The basic unit in meaningful learning—association or principle? *The School Review*, 1967, in press.
- SCANDURA, J. M., & BEHR, M. Prerequisite practice and criterion form in mathematics learning. *Journal of Experimental Education*, 1966, **35**, 54-55.
- SCANDURA, J. M., & ROUGHHEAD, W. G. Conceptual organizers in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, 1967, in press.
- SCANDURA, J. M., WOODWARD, E., & LEE, F. Rule generality and consistency in mathematics learning. *American Educational Research Journal*, 1967, **4**, 303-319.

(Received November 22, 1966)

MEANING AND TRANSFER OF VERBAL LEARNING

JULIUS M. SASSENATH AND GEORGE D. YONGE

University of California, Davis

Synonymity, antonymity, homonymity, and word association "meaning" were studied independent of one another between training and transfer in either the stimulus, response, or both positions in paired-associate lists. In training the stimulus, response, or both, words (for 3 groups of 20 university Ss) were varied in "meaning" in relation to the transfer task in 6 ways: unrelated, synonyms, antonyms, homonyms, associates, and identical. The experiment was a 3 lists by 6 "meanings" design with between Ss for lists and within Ss for meaning. Analyses of variance indicated that there were reliable differences in transfer among lists and among meaning conditions. There was greater transfer for homonyms and associates than for synonyms and antonyms. The discussion focused on preference, response-chaining, and representational mediation hypotheses as well as similarity of meaning in transfer of learning.

The fact that meaningful material is learned more readily than less or nonmeaningful material hardly needs documentation today. Meaningfulness also appears to be an important factor in transfer of learning; however, in this respect the "picture" is less clear. This is true in spite of the large number of research studies on verbal transfer since Gibson (1940) published her studies and analyses in terms of generalization and differentiation. In 1949 Osgood proposed a theory with three laws which integrated the various experimental findings of transfer studies then available and predicted additional transfer phenomena. The result was his well-known transfer surface which holds that the degree of similarity of stimuli, responses, or both, between the training and transfer tasks produces varying degrees of positive or negative transfer effect. Later Osgood (1952) indicated that it was the similarity in meaning (not just phonetic, orthographic or physical similarity) that might be an important factor in mediating positive or negative transfer effects.

Unfortunately, very few studies have been done which have systemat-

ically varied meaning (either synonymity, antonymity, or word associations) in either the stimulus or response elements between paired-associates (PA) lists in training and transfer. Early studies by Haagen (1943), Morgan and Underwood (1950), and Underwood (1951) show that transfer effects increase to the degree that words in the same position between the training and transfer lists are judged similar in meaning. However, later it was realized that in these studies similarity of meaning (synonymity) was confounded with association values between the words employed in the training and transfer tasks. These studies sampled words from Haagen's (1949) lists and he reports that the degree of synonymity and association value correlated .90 for 80 judges rating 400 pairs of adjectives. Thus in the studies by Haagen and Underwood transfer could have been a reflection of either association values, synonymity, or both. Of course, since then Russell and Storms (1955), among others, have reported that association values of words are an important factor in verbal transfer.

In an effort to unsnarl the semantic

from the associative factors, Bastian (1961) designed an experiment in which the same words were used as stimuli in the training and transfer tasks while the response elements were varied from training to transfer. On the other hand, Ryan (1960) did an experiment in which he varied stimulus elements in terms of semantic or associative factors between training and transfer. In both the Ryan and Bastian experiments one-third of the words were high in semantic meaning and low in association value, one-third of the words were high in association value and low in semantic meaning, and one-third were control words which were very low in association value and judged to be very low in similarity of meaning. In both the Ryan and Bastian experiments, that is, whether stimulus or response elements were varied from training to transfer, the results showed positive transfer on most of the measures for both associative and semantic word types as compared with the controls. However, the difference between the associative and semantic word types, though small, favored the associative. Nevertheless, it is difficult to conclude in favor of the word-association hypothesis because in Bastian's study 10 of the 12 word associates and in Ryan's study six of the nine word associates in the transfer list were clearly antonyms of the corresponding stimulus or response element in the training list. Therefore, transfer of learning could have resulted from word associations, antonyms, or both.

In an effort to manipulate systematically both stimulus and response meaning relationships from training to transfer, Wimer (1964) employed adjectives originally used by Osgood (1946). Twenty-five different PA lists in training were constructed in which stimuli or responses were ad-

jectives either identical, similar, unrelated, opposed, or antonyms to adjectives in comparable stimulus or response positions in the transfer list. The design of the experiment was, thus, five stimulus conditions combined factorially with five response conditions. With the exception of the subjects (Ss) with both stimuli and responses identical in training and transfer, none of the other 23 groups showed a statistically reliable difference from the control group (Ss with stimuli and responses unrelated from training to transfer). One difficulty that may have beclouded these results is that stimulus and response elements in the opposed list were not opposite in meaning to the common transfer list but were opposite or really antonyms of the list of words deemed similar, for example, *HARD-soft*. In addition, there was a list of words that were antonyms of the common transfer words, for example, *TENSE-relaxed*.

With the noted difficulties of the previous studies in mind, synonymity, antonymity, and associative meaning still need to be studied independent of one another, but in relation to certain control conditions, and in comparable positions in both stimulus and response elements between training and transfer in one experiment. The present study attempts to do just that.

METHOD

Part I

Materials and procedure. In an effort to select words that were either synonyms, antonyms, or primary associates of one another in the training and transfer tasks, the difficulty was faced of either controlling for or manipulating homonymity as a variable. Because of its possible influence in mediating verbal transfer (especially when *S* responds aloud), homonymity was employed as a variable. In order to develop the materials, a list of 223 pairs of homonyms was constructed. Out of this group, 105 pairs

of homonyms were selected for which a synonym and an antonym could be readily found for at least one of the homonym words in a pair. This was done so that eventually it would be possible to check whether an antonym or synonym would also be a word associate. The 105 pairs of homonyms were then employed in a free-association test. One-half of the pairs of homonyms were put on one form of the test and the other half on another form. In order to control for possible sequence effects in the order of presenting word associates both sets of homonyms were arranged in two different sequences.

Subjects. One of the words of each pair of homonyms, using both sequences, was administered to 114 upperclassmen in educational psychology. Five days later the other word of the pair of homonyms, using both sequences, was administered to the same 114 Ss. Approximately one-half of the Ss received one set of homonyms the first day and the other set the second day, and vice-versa. The frequency of word associations to the homonyms was tabulated on an IBM 7044 computer.

Part II

Materials. Only a limited number of words for the learning lists could be selected whose more frequent associates were *not* a synonym, antonym, or homonym. Table 1 presents the words employed in the training and transfer tasks with notation as to which of three groups of Ss received what PA lists of words in training and transfer. The training and transfer lists consisted of 12 pairs of words. In the training lists either the stimulus, response, or both (for three groups of Ss) were varied in "meaning" in relation to the common transfer list in six ways. Two of the words were unrelated, two were synonyms, two were antonyms, two were homonyms, two were associates, and two were identical words between the training and transfer tasks. Even though each training list was mixed with regard to "meaning," the lists were unmixed or "like" lists according to Underwood (1964). The words employed were limited to two syllables. The frequency of occurrence per 10 million words from Lorge's unpublished count of words from five magazines still being published today ranged from 119 for the two antonyms to 2077 for the two homonyms used in the response elements of the training task. This is a small difference in a count of 10 million words.

TABLE 1
WORDS EMPLOYED AND THEIR RELATIONSHIP IN "MEANING" FOR THE TRAINING TASKS AND COMMON TRANSFER TASK

Types of "meaning"	Stimuli		Responses	
	1	2	3	4
Unrelated	IDLE FAINT	DAWN CAST	sent morn	accept vice
Synonyms	MAIL GRATE	POST GRIND	rows tail	tiers rear
Antonyms	BALL PACKED	CUBE EMPTY	damn hale	bless sick
Homonyms	STAYED GROAN	STAIR GROWN	dear marry	deer merry
Associates	MAIN TOE	STREET FOOT	eight vane	nine weather
Identical	WEAK PEACE	WEAK PEACE	pause soar	pause soar

Note.—The group that had stimuli varied learned paired lists 1-4 in training and 2-4 in transfer; the group that had responses varied learned paired lists 2-3 in training and 2-4 in transfer; the group that had both stimuli and responses varied learned paired lists 1-3 in training and 2-4 in transfer.

Procedure. Training and transfer lists were typed in lower-case letters in the center of 3 by 5 inch white index cards. Standard PA instructions were used prior to the training task. The Ss were run individually by the anticipation method. The Ss were told to say each word out loud. Four sequences of the training and transfer materials were constructed: an initial exposure of the S-R pairs and three sequences for the learning tasks—to prevent serial response learning. S-R pairs in the initial exposure trial were presented at 2-second intervals. In the training and transfer trials stimulus words were presented at S's rate of responding but no word was presented for more than 5 seconds. The correct S-R pair was presented for 2 seconds. In training and transfer Ss performed to a criterion of two successive trials without error. (The number of trials for each condition of "meaning" was constant since a repeated-measures design was employed.) Immediately after reaching criterion on the training task, S was told "now I will show you another set of cards. The instructions are the same as for the last set."

Design and subjects. The experiment was a three lists (stimuli, responses, or both varied from training to transfer) by six types of "meaning" design; with a between-Ss for lists and a within-Ss for meaning. Sixty Ss were assigned alternately to either the stimulus, response, or stimulus and response conditions in training as they appeared at the laboratory. The Ss were

TABLE 2
MEAN NUMBER OF ERRORS IN TRAINING

Types of "meaning"	Lists		
	Stimulus	Response	Stimulus and response
Unrelated	5.90	6.85	6.30
Synonym	6.00	4.80	4.75
Antonym	4.05	5.45	4.45
Homonym	5.05	4.95	5.05
Associate	3.15	4.10	4.90
Identical	5.25	5.00	5.60
Total	29.40	31.15	31.05

volunteers from classes in educational psychology and had not participated in Part I of the study.

RESULTS

Training

The mean number of errors over all the trials in training are presented in Table 2. An analysis of variance for repeated measures indicates that there are no reliable differences in performance among the three lists ($F < 1.0$; $df = 2/57$). However, there are reliable differences among the types of meaning ($F = 5.37$; $df = 5/285$; $p < .01$), but the interaction of Lists $\times$ Meaning ($F = 1.20$; $df = 10/285$) is not significant. In order to locate where the differences among the meaning conditions existed, Scheffé's confidence interval procedure was applied to the data for each of the three lists (groups of Ss) separately; using the within Ss mean square error term from a one-way analysis of variance for repeated measures with $\alpha = .01$. It was found, only in the stimulus list, that the unrelated and synonym conditions had reliably more errors than the association condition. In the response and stimulus and response lists there were no differences among the means. Thus the words in the associate condition for the stimulus list only were easier to learn than those in the unre-

lated and synonym conditions, even though a pilot study showed no differences.

Transfer

In view of the significant difference among meaning conditions for the stimulus list in training, rank-order correlations were computed between over-all errors in training and transfer for each cell in the six meaning by three lists conditions to ascertain whether analyses of covariance should be employed. The correlations ranged from .40 to $-.31$ with a median value of .10. Since only two of the 18 correlations were significantly different from zero at the .05 level, it was decided that analyses of covariance were unwarranted.

First trial. Table 3 presents the mean number of errors on the first trial in transfer. An analysis of variance showed that the differences among the three lists were significant ($F = 6.20$; $df = 2/57$; $p < .01$). The total errors on the first list were what could be expected from transfer theory; where only stimuli are varied there are fewer errors than where only responses are varied, and where responses are varied but are similar there are fewer errors than where both are varied. Also there was a reliable difference among the meaning condi-

TABLE 3
MEAN NUMBER OF ERRORS ON THE FIRST TRIAL AND ALL TRIALS IN TRANSFER

Types of "meaning"	First trial			All trials		
	Stimulus	Response	Stimulus and response	Stimulus	Response	Stimulus and response
Unrelated	.95	1.55	1.40	1.55	2.55	2.15
Synonym	.60	.90	1.20	1.00	1.45	1.85
Antonym	.75	1.05	1.50	1.05	2.15	2.00
Homonym	.35	.10	.50	.45	.35	.55
Associate	.35	.20	.50	.40	.45	.70
Identical	.35	.20	.05	.70	.45	.25
Total	3.35	4.00	5.15	5.15	7.40	7.50

tions ($F = 12.97$; $df = 5/285$; $p < .001$) and the Lists $\times$ Meaning interaction was significant ($F = 3.33$; $df = 10/285$; $p < .001$).

In order to obtain an estimate of the degree of positive and negative transfer on the first trial each mean can be compared with the unrelated S + R mean in Table 3 which is the A-B, C-D control for learning-to-learn and "warm-up" effects. Without going into all the comparisons, it can be readily seen that when the words in training are homonyms, associates, or, of course, identical to their counterpart in the stimulus, response, or stimulus and response conditions on the transfer task, there is considerable positive transfer effect.

A one-way analysis of variance with repeated measures was applied to the stimulus list, and the meaning condition was significant ($F = 3.91$; $df = 5/95$; $p < .01$). However, when Scheffé's confidence-interval procedure was used, with the appropriate mean square error, none of the mean differences were reliable. It appears that Scheffé's test is a conservative one. Since Bastian and Ryan found that the mean errors were less for the associate than the semantic conditions, a t test for correlated means with a one-tailed test was used. Indeed, the difference was significant ($t = 2.60$; $df = 19$; $p < .05$).

For the response list an analysis of variance for repeated measures was, again, employed, and the meaning condition was significant ($F = 27.80$; $df = 5/95$; $p < .001$). Scheffé's procedure showed that the synonym, homonym, associate, and identical conditions were different from the unrelated. In addition, the homonyms, associates, and identicals are different from the antonyms as well as the synonyms.

For the stimulus + response list

the analysis of variance for meaning was, again, significant ($F = 17.32$; $df = 5/95$; $p < .001$). Scheffé's procedure indicated that the homonym, associate and identical conditions were reliably different from both the unrelated and antonym conditions. Furthermore, the identical was different from the synonym condition.

All trials. Table 3 also presents the mean number of errors for all the trials in transfer. Employing the same analyses to these data as for the first trial, it was found that there were reliable differences for lists ($F = 3.40$; $df = 2/57$; $p < .05$). These results indicate that there were fewer errors on the stimulus than the other two lists. Also, the meaning conditions ($F = 28.82$; $df = 5/285$; $p < .001$) and the interaction of List $\times$ Meaning was significant ($F = 2.12$; $df = 10/285$; $p < .05$). An indication of the amount and direction of specific transfer effects can be had by comparing each mean with the A-B, C-D control for general transfer effects. In general, it can be seen that when the words in training are homonyms, associates, or identical to the corresponding words in the stimulus, response, or stimulus and response conditions in the transfer task there is a great deal of positive transfer.

For the stimulus condition a repeated measures analysis of variance indicated that meaning was significant ($F = 3.75$; $df = 5/95$; $p < .01$). When Scheffé's procedure was employed, however, none of the mean differences was reliable. Again, since Bastian and Ryan found that the mean errors were larger for the semantic than the associates, a one-tailed t test for correlated groups was computed and found to be significant ($t = 2.60$; $df = 19$; $p < .01$).

An analysis of variance with repeated measures was applied to the

response list, and there were reliable differences among the means ($F = 13.92$; $df = 5/95$; $p < .001$). Scheffé's procedure showed that the homonyms, associates, and identical words were reliably different from the unrelated and antonym words.

For the stimulus + response list the analysis of variance showed that the meaning conditions were again significant ($F = 14.39$; $df = 5/95$; $p < .001$). Scheffé's procedure indicated that the homonyms, associates, and identical words were reliably different from the unrelated and antonym words. In addition, the homonyms and identicals were different from the synonyms.

DISCUSSION

As far as "meaning" is concerned, the results of the experiment lead to the conclusion that in paired-associate learning when the relationship between training and transfer words are associates or homonyms there is considerably more transfer of learning than when the relationship is one of antonyms or synonyms. Furthermore, the associates and homonyms appear to have the same effect as practicing the identical words from training to transfer. This seems to be the case whether the stimulus, response, or both are varied. The fact that there was greater transfer for words with associative meaning than synonym (semantic) meaning where either stimuli or responses are varied confirms the previous findings by Bastian (1961) and Ryan (1960). However, since Bastian and Ryan confounded word association with antonymity the results of the present experiment indicate that they were correct in inferring that there was more transfer for words having associative connections than semantic meaning. This is borne out by the present study in which there

were no reliable differences in transfer between the synonym and antonym conditions, but there were reliable differences between the antonym and associate conditions.

Of course, the results obtained in this experiment may not occur when homogeneous or unmixed lists for meaning are employed, that is, where six separate groups learn six separate lists followed by a common transfer task. Underwood (1964) suggests that in mixed lists *S* may have a "preference" for starting with certain items and will learn these first. Whereas a list homogeneous for the same kind of words may not be learned faster than a list homogeneous for other kinds of words. There was no "preference" for any of the types of "meaning" words in the three training tasks, with the exception of associates over the unrelated and synonyms words in the stimulus-varied list. Therefore, any preference in the transfer task would have to be based upon meaning relationships between training and transfer. One interpretation is that with mixed lists differential transfer may be based upon preference, particularly preference for associates and homonyms.

Perhaps the results of the present experiment also can be interpreted as favoring the associative, response chaining, mediation theory espoused by the Minnesota researchers (Jenkins, 1963; Russell & Storms, 1955) who have employed free-word-association procedures from which mediation processes can be inferred in transfer. On the other hand, some may wish to interpret the results as not favoring the representational mediation theory of Osgood (1952), who has advocated the use of the semantic differential for the assessment of mediators. However, the semantic differential is probably not measuring semantic (syno-

nym, antonym) or lexical meaning but connotative meaning. Therefore, since there was less transfer for the synonym and antonym than the word-association conditions in the present experiment, these results do not appear to be contradictory to Osgood's representational mediation theory. As a matter of fact, some psychologists (Bousfield, 1961; DeBurger & Donahoe, 1965; Noble, 1963) have indicated that associative meaning and connotative meaning, as measured on the semantic differential, are highly correlated, at least for some words. Osgood (1961) has presented evidence that *S*'s marks on the scales of the semantic differential were not made only on the basis of word-association tendencies.

A curious finding in the present study is that there was more transfer for the homonym than the synonym condition. This result is contrary to those obtained by Razran (1939) and Riess (1946), who found more generalization among synonyms than homonyms for adult *Ss*. Perhaps some of this discrepancy can be attributed to the methods employed. Razran and Riess used a classical, semantic-conditioning paradigm whereas the present study employed an instrumental learning paradigm. Further, the fact that *Ss* pronounced all words may have served to make the homonyms function somewhat like identical words. Also, different words were employed.

With regard to Osgood's (1949) transfer model and the question of similarity or, particularly, the opposed relationship between training and transfer words, a logically and semantically confusing point should be clarified. Some psychologists (e.g., Wimer, 1964) have apparently thought that semantic similarity could be scaled from identical, to similar, to

unrelated, to opposed, to antonyms. Words judged to be opposed or antonyms of other words must be semantically related, otherwise opposition or antonymity could not be ascertained. Consequently, opposite or antonym words must fall on a semantic scale somewhere between being identical and unrelated rather than at points beyond that of unrelatedness. If they were semantically beyond the point of unrelatedness, negative transfer should occur. However, instead of showing negative or even no transfer, when antonyms were varied in either the stimulus, response, or both conditions in the present experiment there was some positive transfer when compared to the corresponding unrelated condition for all three groups of *Ss*. These results place antonyms on a scale between identical and unrelated words.

The results obtained under the conditions of the present study should not be inferred to suggest that educators should stress the teaching of homonym words or mere associations to words as opposed to the semantic aspects of words. On the contrary, the semantic aspects of words are probably far more important, educationally speaking. Synonyms, antonyms, and other semantic aspects of words are often learned in a context (syntax) and have a transfer value in a context. Whereas synonyms and antonyms out of context, as in the present experiment, have some transfer value, it is far less than word associates and homonyms out of context.

REFERENCES

- BASTIAN, J. Associative factors in verbal transfer. *Journal of Experimental Psychology*, 1961, **62**, 70-79.
- BOUSFIELD, W. A. The problem of meaning in verbal learning. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 81-91.

- DEBURGER, R. A. & DONAHOE, J. W. Relationship between the meanings of verbal stimuli and their associative responses. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 25-31.
- GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, 1940, **47**, 196-229.
- HAAGEN, C. H. Learning and retention as a function of the synonymy of original and interpolated tasks. Unpublished doctoral dissertation, State University of Iowa, 1943.
- HAAGEN, C. H. Synonymy, vividness, familiarity and association value ratings of 400 pairs of common adjectives. *Journal of Psychology*, 1949, **27**, 453-463.
- JENKINS, J. J. Mediated associations: Paradigms and situations. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963. Pp. 210-245.
- MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *Journal of Experimental Psychology*, 1950, **40**, 592-603.
- NOBLE, C. E. Meaningfulness and familiarity. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning*. McGraw-Hill, 1963. Pp. 76-119.
- OSGOOD, C. E. Meaningful similarity and interference in learning. *Journal of Experimental Psychology*, 1946, **36**, 277-301.
- OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychological Review*, 1949, **56**, 132-143.
- OSGOOD, C. E. The nature and measurement of meaning. *Psychological Bulletin*, 1952, **49**, 197-237.
- OSGOOD, C. E. Comments on Professor Bousfield's paper. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 91-106.
- RAZRAN, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89-90.
- RIESS, B. F. Genetic changes in semantic conditioning. *Journal of Experimental Psychology*, 1946, **36**, 143-152.
- RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *Journal of Experimental Psychology*, 1955, **49**, 287-293.
- RYAN, J. J. Comparison of verbal response transfer mediated by meaningfully similar and associated stimuli. *Journal of Experimental Psychology*, 1960, **60**, 408-415.
- UNDERWOOD, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first list learning. *Journal of Experimental Psychology*, 1951, **42**, 44-53.
- UNDERWOOD, B. J. The representativeness of rote learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964. Pp. 47-78.
- WIMER, R. Osgood's transfer surface: Extension and test. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 274-279.

(Received December 2, 1966)

READING RATE AND IMMEDIATE VERSUS DELAYED RETENTION

STANTON P. THALBERG
University of Washington

The purpose of this study was to determine whether the relationship between reading rate and retention is invariable over time. 176 college Ss were randomly assigned to 1 of 2 conditions of retention. Ss in the immediate treatment (IR) were tested immediately after reading a 1500-word passage. Ss in the delayed treatment (DR) read the same passage but were tested 24 hrs. later. Ss were subdivided within treatments into fast-, average-, and slow-rate groups. Results indicated that under IR conditions slow readers retained significantly more than both average and fast readers. Under conditions of DR, retention differences between rate-groups disappeared. Implications are that while more efficient readers remember fewer of the details in a message immediately following the reading than do their slower counterparts, these details extinguish for both groups equally within 24 hrs.

Studies concerned with the relationship between reading rate and comprehension date back to Romanes' (1884) research conducted over eight decades ago. Time has apparently neither detracted from the importance of considering this relationship nor relegated it to an obscure position in educational research, despite the relative tapering off of research bulk during the 1960s. The continuance of investigation is due at least in part to the divergent results already obtained. The relationships between rate and comprehension, as reported in the professional literature, range from moderately negative ($r = -.47$) to highly positive ($r = .98$) with the great majority of studies yielding a negligible or slightly positive correlation coefficient.

Numerous explanations for these discrepant results have been postulated and systematically included into the body of professional knowledge. These explanations have included the nature of the reading material (Thurstone, 1944), the difficulty of the reading material (Tinker, 1939), the purpose and set of the reader (Carlson, 1949; Shores & Husbands, 1950), the intellectual

level of the reader (Carlson, 1949), and the techniques employed in measuring both the reading rate and level of comprehension (Preston & Botel, 1951; Stroud, 1942, 1956; Thalberg, 1959). The impact of these studies has resulted in a tempering of judgment and qualifying of pronouncements regarding the predictability of either rate or comprehension from mere knowledge of the other.

Yet, despite the impressive accumulation of research, it is neither exhaustive nor plethoric. For example, all preceding research has assessed comprehension and retention immediately following the reading, the apparent premise being that what is true of the relationship under conditions of immediate recall holds equally well if an extended period of time intervenes between reading and the measurement of retention.

Early investigators of the relationship between rate of learning and retention tended toward a practice-equal learning-variable paradigm (Gates, 1918; Gillette, 1936; Thorndike, 1910) with the understandable result that fast learners retained more than slow learners.

More recent research, however, has

tended toward a learning-constant/practice-variable model in which learning to a given criterion is the independent variable and the number of trials to achieve this level of performance is the dependent variable (Stroud & Schoer, 1959; Underwood, 1954). Under these conditions, the relationship between rate of learning and retention is no more than slight.

Since the practice of encouraging students to increase their reading rate persists, it was appropriate to view the rapid-reading phenomenon from a slightly different vantage point.

The current emphasis upon rate encourages students to read more rapidly and promises no loss in immediate retention; however, this increment in speed may facilitate, inhibit, or bear no relationship to long-term retention. The sole provision of immediate rewards for correct responses and the exclusion of delayed reinforcements may, in fact, be conditioning the students to the persuasion that no additional benefits will accrue if the information is retained beyond the immediate test period.

This study, then, seeks information relative to the impact of reading rate on immediate and delayed retention. In brief, answers to two questions were sought: Does reading rate differentially affect the magnitude of immediate and delayed retention? Is the relationship between rate and comprehension invariable and therefore independent of the time interval between reading and the measure of retention?

METHOD

The experimental subjects (Ss) consisted of 176 college undergraduates enrolled in 13 sections of freshman rhetoric (freshman English) at the University of Iowa. These participants were randomly assigned by in-

tact class sections to one of two experimental treatment groups. The Ss were further subdivided within treatments into three rate levels. Certain Ss (not included in the 176) were excluded from the study for the following reasons: 19 for whom American College Test (ACT) scores were not available, 11 who had incomplete response data, and 9 who were eliminated randomly to effect proportionality between cells of the experimental design.

The Traxler High School Reading Test, Form B, Part 1 (Traxler, 1939) was employed as the measure of both reading rate and retention. Despite the fact that the test was not primarily intended for use at the college level, it was utilized in this investigation for the following reasons. First, it consists of a reading passage of approximately 1500 words, a length sufficient for a relatively stable measure of reading rate. Second, although the test deals with a relatively sophisticated and technical topic (the historical development of modern-day calendars), its readability is at a standard level, that is, written between an eighth and ninth-grade level as determined by both the Dale-Chall (1948) readability formula and the Flesch Reading Ease Score (1948). Third, the test provides a means for measuring both rate and comprehension over the same material. Fourth, a larger number of questions (20) is provided than is true of other comparable tests, thereby providing a more stable and reliable measure of reading comprehension. Finally, the test is amenable to an amount-constant assessment of rate rather than a time-constant measure.

The directions for administering the test were adapted from the modified time-constant measure of rate recommended by the publishers to an amount-constant approach in order to assure a more reliable and representative estimate of rate and a more factual measure of comprehension. The following specific directions were read to Ss:

This is a test of your overall reading ability. The test contains a story for you to read and some exercises for you to work after you have read the story. Read the article through once for the primary purposes of learning and retaining as much as possible of its content. You will be asked questions over the article. Read rapidly but stress your understanding. Remember, you cannot work the exercises over the content unless you know and remember what you have read.

In determining their rate of reading, Ss were told that immediately upon finishing

the reading they were to record the number of elapsed seconds—taken from the chalk board—on their answer sheets. Ss in the Immediate Recall (IR) treatment-group were instructed to answer the comprehension questions immediately after recording their reading time. Ss in the Delayed Retention (DR) treatment-group were instructed to close their test booklets, record their reading time, and await further instructions. Within this latter group (after all Ss had completed the reading) the materials were collected and the participants were dismissed. The following day the materials were redistributed and each S was required to respond to the questions without further reference to the article.

Within the two treatment groups Ss were subdivided into three levels based upon their performance on the rate aspect of the test. A high reading group was composed of Ss reading 300 or more words per minute, an average reading group contained Ss reading between 240 and 299 words per minute, and a low-rate group was composed of Ss reading 239 or fewer words per minute. The final constitution of these three groups by treatments is reported in Table 1. None of the differences between mean rates within levels was significant at the .05 level.

ACT scores were available on each of the 176 Ss. The means range from 45.41 to 51.16, and none of the differences between subgroups was statistically significant at the .05 level. Consequently, any possible impact of verbal ability upon the criterion measures was argued to be either nonexistent or minimal.

RESULTS AND DISCUSSION

Mean scores and standard deviations for each of the three experimental subgroups on both measures of retention are reported in Table 2.

TABLE 1
TABLE OF RATE MEANS BY LEVELS AND
RETENTION CONDITIONS

	<i>N</i>	Mean	<i>SD</i>
Immediate retention	(80)		
Fast rate	25	321.36	55.95
Average rate	25	252.76	9.42
Slow rate	30	207.90	20.62
Delayed retention	(96)		
Fast rate	30	340.86	66.35
Average rate	30	255.40	11.44
Slow rate	36	215.44	17.35

TABLE 2
RETENTION MEANS AND STANDARD DEVIATIONS BY RATE LEVELS AND RETENTION CONDITIONS

	<i>N</i>	Mean	<i>SD</i>
Immediate retention	(80)		
Fast rate	25	11.28	2.85
Average rate	25	12.56	2.04
Slow rate	30	13.66	1.99
Delayed retention	(96)		
Fast rate	30	10.40	2.47
Average rate	30	9.83	1.98
Slow rate	36	10.91	2.20

Retention scores for the comprehension task were submitted to a two-dimensional (Rate of Reading $\times$ Retention Condition) treatment-by-levels analysis of variance (Lindquist, 1953). The main effects of retention and rate were both significant, as was the interaction effect. IR scores were significantly superior to those of DR, the former condition yielding a mean of 12.57 and the latter a mean of 10.41 ($F = 38.35$, $p < .001$). These findings are consistent with the principles of retroactive and proactive inhibition.

Of greater interest than the superiority of the IR condition over DR is that when collapsed across treatments the three rate levels are distinguishable by level of comprehension ($F = 6.08$, $p < .01$). The mean retention score of the high-rate group was 10.80, the average-rate group 11.07, and the low-rate group 12.16. The consistency of the inverse relationship was somewhat surprising in view of the voluminous studies which suggest that the relationship between rate and comprehension is virtually nil. In this regard the low-rate group was found to be significantly higher than both the average and the high groups ($t = -2.59$, $df = 1/64$; $t = -3.23$, $df = 1/53$; respectively) at the .01 level.

Because of the significant interaction effect ($F = 3.03$, $p < .05$), a

TABLE 3
CORRELATION MATRIX FOR ABILITY, RATE,
AND RETENTION

	ACT	Rate	Comp
ACT		-.034	.422*
Rate	-.042		-.385*
Comp	.288*	-.193*	

Note.—Correlations above diagonal are immediate retention ($N = 80$); below are delayed ($N = 96$).

* $p < .05$.

subsequent analysis of treatments-within-rate-levels employing a simple randomized design was conducted and indicated a significant treatment effect ($p < .01$) within the average- and low-rate groups. No statistically significant differences were forthcoming within the high group. In other words, average and slow readers, while ostensibly comprehending more immediately after reading than the fast readers, showed a relatively high rate of deterioration of their comprehension over time. On the other hand, the rapid readers, although observably less than other comparison groups initially, found their initial level of comprehension virtually intact over time: They lost only .88 questions between immediate and delayed trials, whereas the average group lost 2.73 and the low group 2.75 questions.

The experimental design was broken down further to allow for an analysis of differences between rate groups within retention conditions. Within the IR treatment the low-rate group was significantly higher in understanding than both the high- ($F = 13.30$, $p < .01$) and the average-rate groups ($F = 4.12$, $p < .05$). Under the DR treatment, however, the

low-rate group maintained its superiority over only the average group ($F = 4.27$, $p < .05$), the difference between the low and the high groups apparently dissipating with time.

Two alternatives present themselves as possible explanations of the finding that retention differences between readers of varying rates, though clearly favoring the slow readers immediately following the period of acquisition, wash out with time.

On the one hand, it can be postulated that the slow reader receives more vivid sensory impressions of the material because of the dotting nature of his reading, which, in turn, allows him to accurately feedback the material immediately. However, the advantage gained initially through this increased clarity of the sensory impressions rapidly diminishes—probably as a result of retroactive inhibition or disuse—so that the temporal enhancement almost totally disappears after twenty-four hours or less. No theory has yet been advanced which postulates that vivid sensory impressions simultaneously lead an individual toward more concrete and functional initial understandings and to fewer, less effective understandings with the passage of time. Indeed, clinical experiences suggest the opposite conclusion: The more vivid a sensory experience, the longer it is retained.

On the other hand, it may be postulated that the slow reader—by virtue of more deliberate attempts to understand—is more attentive to minor points and details within the message, while the faster reader is more concerned with the organization of these details into meaningful generalizations. After a 24-hour intervening period, however, these details deteriorate for all readers (regardless of

rate) and only the more generalized concepts remain. These generalizations apparently are well-retained by readers of all rates.

Although no analysis was made of the responses of Ss of various reading rates to different types of comprehension questions (detailed or generalized), the investigator did informally examine Ss' responses with this in mind. The Ss with a slow or average rate tended to perform better than fast readers on questions involving details during IR; yet during the DR condition all three classes of Ss tended more toward correct responses on the general-type questions and toward more incorrect responses on questions dealing with details.

Table 3 presents the correlation matrix pertinent to the constancy over time of the relationship between rate and comprehension. Although the coefficients of $-.385$ under the immediate condition and $-.193$ under the delayed appeared to differ sufficiently for statistical significance, analysis by the z' Transformation (Edwards, 1954) indicated that the discrepancy was not large enough to reach significance ($p > .15$). Consequently, one is led to the conclusion that the relationship between rate and comprehension is relatively invariant.

The relationship between ability (as measured by the ACT) and rate and the relationship between ability and comprehension under both retention conditions were also investigated. As can be seen in Table 3, no apparent relationship between ACT and rate under either condition was noted, the correlation coefficients being $-.034$ and $-.042$ under immediate and delayed trials, respectively. The relationship between ACT and comprehension was sta-

tistically significant ($p < .05$), the coefficients being $.422$ for immediate and $.388$ for delayed retention. In other words, these relationships suggest that reading rate is independent of ability and is apparently more of a mechanical or perceptual nature than of an intellectual one. Conversely, however, ability to comprehend and retain verbal material is, at least in part, a function of intellectual ability.

If one can generalize from the results of this investigation to their educational implications, the relative efficiency of the faster reader over that of the slow or average reader maintains itself over time. That is to say, although the more efficient reader may not comprehend as much minutiae immediately following his reading than either the average or the slower readers, he will in the long run have as much functional information available to him from his reading as either of the other levels. Further, the less efficient reader appears to forget more of what he reads over time than the more efficient reader. This seems to imply that reading improvement courses which stress rate and comprehension equally will not detrimentally influence the retention of their clientele.

REFERENCES

- CARLSON, T. R. The relationship between speed and accuracy of comprehension. *Journal of Educational Research*, 1949, **42**, 500-512.
- DALE, E., & CHALL, J. F. A formula for predicting readability. *Educational Research Bulletin*. Columbus, Ohio: Ohio State University, 1948, **27**, 11-20.
- EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.
- FLESCH, R. A new readability yardstick. *Journal of Applied Psychology*, 1948, **32**, 221-233.
- GATES, A. I. Correlations of immediate and

- delayed recall. *Journal of Educational Psychology*, 1918, **9**, 489-496.
- GILLETTE, A. L. Learning and retention, a comparison of three experimental procedures. *Archives of Psychology*, 1936, **28**, No. 198.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin, 1953.
- PRESTON, R. C., & BOTEL, M. Reading comprehension tested under timed and untimed conditions. *School and Society*, 1951, **74**, 71.
- ROMANES, G. J. *Mental evolution in animals*. New York: Appleton, 1884.
- SHORES, J. H., & HUSBANDS, K. L. Are fast readers the best readers? *Elementary English*, 1950, **27**, 52-57.
- STROUD, J. B. A critical note on reading. *Psychological Bulletin*, 1942, **39**, 173-178.
- STROUD, J. B. *Psychology in education*. New York: Longmans, Green, 1956. P. 160.
- STROUD, J. B., & SCHOER, L. Individual differences in memory. *Journal of Educational Psychology*, 1959, **50**, 285-292.
- THALBERG, S. P. Reading rate and comprehension in a college reading program. Unpublished master's thesis, University of Iowa, 1959.
- THORNDIKE, E. L. The relation between memory for words and memory for numbers, and the relation between memory over short and memory over long intervals. *American Journal of Psychology*, 1910, **21**, 487-488.
- THURSTONE, L. L. *A factorial study of perception*. Chicago: University of Chicago Press, 1944.
- TINKER, M. A. Speed versus comprehension in reading as affected by level of difficulty. *Journal of Educational Psychology*, 1939, **30**, 81-94.
- TRAXLER, A. J. *Traxler high school reading test*. Indianapolis: Bobbs-Merrill, 1939.
- UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychological Bulletin*, 1954, **51**, 276-282.

(Received December 6, 1966)

PERCEPTION OF DEVIANT BEHAVIOR: A STUDY OF CHEATING

JAMES Q. KNOWLTON
Indiana University

AND

LEO A. HAMERLYNCK¹
University of Oregon

2 investigations of "cheating" were conducted; one at a small liberal arts college ($N = 165$), and the other at a large metropolitan university ($N = 533$). Anonymous questionnaires were administered to representative samples of the 2 student bodies, and relationships between the extent of admitted cheating behavior, estimates of the amount of cheating within the college or university, and attitudes toward cheating were determined. The group of individuals who classified themselves as cheaters tended to give higher estimates of the extent of cheating by others than did noncheaters. These same individuals tended to be less condemning of cheating, and to explain cheating as being due to environmental pressures. The noncheating group described the cheater as an individual having a basic personality defect. (Reservations concerning this conclusion were made.) Of 3 classes of situations thought conducive to cheating—adverse physical conditions in the classroom, inadequate tests and testing procedures, and instructor failings—instructor failings were considered most responsible for cheating by both the cheater and noncheater groups. (Tested only on School A students.) Differences between cheaters and noncheaters were observed on both personal and demographic levels. Alternative explanations of the findings were considered.

Why do students cheat on examinations? To gain a better grade? Perhaps. But not many students put it quite this way. Cheating is thought by a sizeable number of students (12% of one student body), to be just "playing the game with the professor." However, the authors have found student responses that ranged all the way from extreme condemnation of the practice ("cheaters should be put in jail as common thieves") to views expressing the belief that cheating is praiseworthy ("successful cheating gives one a pride of accomplishment").

This report describes two studies

¹ Many persons gave unstintingly of their time in completing various phases of the two studies reported here. The late Franklin Fearing served as a constant and thorough critic. Stanley Kegel helped in the collection and coding of data, and Bernice Eisman in the analysis. At both schools, service organizations truly proved to be of service. The administrations of both schools gave encouragement, while following a strictly "hands-off" policy.

of cheating. Both studies were investigated by administration, faculty, and student groups, suddenly alarmed at the "cheating situation" on their campuses. The student-faculty-administration promoters of these studies had no theoretical interests, but the present writers tried to reap a theoretical by-product as a dividend for the work involved.

The hypotheses for both studies were based on what has been called "social-perceptual theory." Else Frenkel-Brunswik (1951), for example, in a discussion of the kinds of statements people make in reference to their own shortcomings, points out that to project one's shortcomings onto the environment is a comfortable way to deal with what would otherwise be unflattering self-thoughts, for one thereby remains unperceptive of one's failings. As Frenkel-Brunswik describes it, "we do not always see ourselves as we are, but instead preceive the environment in

terms of our own needs. Self-perception and perception of the environment actually merge in the service of these needs [1951, p. 379]."

If to label oneself a "cheater" is an unflattering self-thought, the kinds of dynamics Frenkel-Brunswik speaks of would be predicted to come into play. This was the orientation that guided the derivation of the hypotheses for test. However, the findings to be reported can as well be explained within the bounds of another theoretical framework; and since this unanticipated mode of explanation is postdictive, discussion of it is saved for the last part of the paper.

The two studies here described are essentially replications of one another. Replication was thought to be of more than usual value because of wide differences between the two schools studied. School A is a small, 4-year, liberal arts college in a small town in a sparsely populated area. School B is a large metropolitan university located in a densely populated area and has a heavy graduate program. Because of these differences between the schools, the generality of any results supported by both studies should be broad.

Throughout, it has been assumed that honesty is a continuously distributed variable that is defined in terms of the situation as much as by the specific act. But in order to avoid an involved style, the authors will use the terms "cheater" and "noncheater," even though to do so implies both a nonexistent dichotomy and a degree of absoluteness that is unwarranted.

METHOD

The School A Study

A questionnaire (see ADI document A)* was developed to elicit student judgments

* Material supplementary to this article

about the cheating behavior of themselves and others, their attitudes toward cheating, their opinions about the "cheating situation" at School A, and personal data.

The sample was 10% of the regular undergraduate, full-time student population. The seventh (chance determined), seventeenth, etc. students in the registrar's alphabetical listing were chosen, and the necessary locating information recorded. These students received a letter requesting that they present themselves at a central polling station to give their opinions on the "cheating situation" at School A. The student was promised complete anonymity by direct statements in the letter, in the student newspaper, in the questionnaire, and by suggestion. For example, his letter was signed by the student body president, indicating student backing; the polls were manned by members of a school service organization, and no faculty or administrative person had contact with the respondents. At the central polling station, the student was told that when completed he was to fold his questionnaire and drop it in the locked ballot box provided. He completed the questionnaire at a screened table in comfort and privacy.

During the first week 40% of the sample responded. Nonrespondents were then telephoned or sent a post card, and another 30% responded. Finally, members of the service organization were given the names of the remaining nonrespondents along with their class schedules. These persons were personally contacted and were requested to fill out the "cheating questionnaire." With this the final N became 165, or 89% of the sample of 186.

Checks on sample representativeness (in terms of the sample proportion of freshmen, sophomores, etc., compared with these proportions as recorded by the registrar), showed all differences between sample and universe to be insignificant (all *p* values were in the .90's).

The School B Study

This questionnaire is available as ADI document B. Five percent of the full time

has been deposited with the American Documentation Institute. Order Document No. 9632 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance \$3.75 for microfilm or \$2.50 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

undergraduate and graduate student body made up the sample. After three contacts, when necessary, and including a personal contact the third time, a sample of 533 (a 90% return) was achieved. Representativeness was checked as in the first study, and very close universe/sample proportions were noted. Time of reporting—that is, after one, two, or three requests—was not significantly related to whether or not the student admitted cheating, or expressed condemning or condoning attitudes toward cheating ($p > .20$ and $p > .30$ respectively).

RESULTS

Hypothesis 1

Individuals who deviate from group prescribed norms will differ significantly from those who do not deviate in giving higher estimates of the number of individuals who deviate as they do. In the present instance, the cheater will perceive more cheating going on about him than will the non-cheater.

School A study. One variable was the subject's (*S*'s) reply to a question asking: "In your opinion, about what percent of [School A] students cheat fairly regularly?" Since *S* could have no objective basis for his judgment, his response was regarded as reflective of his personal outlook. As expected of a projective item, this one yielded an extremely wide range of responses (from less than 10% to over 90%; median = 23%).

The second variable was *S*'s statement as to the number of situations in which he had cheated. These situations described "rationalizations" for cheating. An example is: "I have cheated when so many others were doing it that one had to do so in order to have a fair chance in competing for grades." The *S* was instructed to indicate those statements that applied to himself. There were eight rationalizations, and "cheating scores" could therefore range between zero (cheating denied in all

situations), and eight (the respondent admitted cheating in every situation described).

The School A data tend to confirm Hypothesis 1: contingency coefficient (C) = .36, $N = 161$, $p < .02$, 4×5 table. (See ADI table A.)³

School B study. The projection variable in the second study was measured in terms of responses to this question:

In your opinion about what proportion of students at [School B] frequently cheat? (Check one).

- (a) practically all of them
- (b) most of them (i.e., about 75% of them)
- (c) about half of them
- (d) a few of them (i.e., about 25% of them)
- (e) practically none of them

A "high-perceiver" group, a "low-perceiver" group, and a "medium-perceiver" group were defined as those who checked (a) or (b), (d) or (e), or (c), respectively.

The *S*s were also asked to indicate their cheating behavior in terms of a three-fold breakdown: "active cheating," "passive cheating," and "non-cheating." The questionnaire defined active cheating as "cheating in which the person benefits by such an act," (e.g., receiving unauthorized information). Passive cheating was defined as "cheating in which another

³In lieu of tabularly presented data, the authors report the contingency coefficient (C), and also N , p , and the number of categories (Row $\times$ Column) in the table under consideration. This last, of course, determines the theoretical upper limit of the contingency coefficient; a value ranging from the mid .70s through the .80s for tables of the size employed in this study.

The median value of a sample of 112 correlation ratios (η) derived from studies reported in recent APA journals was .42, range .05–.92 (Dunnette, 1966). These correlation ratios (derived from the t and F values reported by the authors involved) have a theoretical upper limit of 1.0.

individual benefits by the act," (e.g., giving unauthorized information to another). Those indicating they had cheated only actively, or both actively and passively, were designated "active cheaters" ($N = 252$). Those who stated that they had not cheated at School B were designated "non-cheaters" ($N = 185$). Except for nine nonrespondents to the cheating question, the remainder ($N = 87$), claimed to have cheated only passively.

The hypothesis that active, passive, and noncheaters perceive significantly different amounts of cheating in the university was tested by means of chi square. The hypothesis tended to be confirmed with $C = .29$, $N = 524$, $p < .001$, 3×3 table. (See ADI Table B.)

Hypothesis 2

Individuals who deviate from group norms will be less condemnatory of this behavior than those who do not deviate. The cheater, in condemning cheating less strongly than the noncheater, thereby perceives others to be less condemning of him and he feels less subject to conformity-producing pressures.

School A study. The S 's cheating score was compared with his attitude score, which was derived in the usual manner for Likert scales, and so that those who condoned cheating most earned the highest scores. The School A data tend to confirm Hypothesis 2 with $C = .42$, $N = 165$, $p < .001$, 3×5 table (ADI Table C).

School B study. The 10 attitude statements used in the School B study were not scaled, and composite attitude scores were therefore unavailable. Consequently, Hypothesis 2 was tested by separately comparing the different cheating group responses to

the following four items, one at a time.

1. "Cheating is basically as immoral as stealing."

2. "As an ideal, honesty in examinations deserves my support, but students are human and cannot be expected at present to attain this ideal."

3. "Cheating reveals a basic defect in the character of the student."

4. "Cheating is basically wrong but is sometimes necessary when too much emphasis is placed upon grades."

These particular items were selected as potentially good discriminators because the percentage of all respondents in agreement with them ranged closely about the 50% mark (46%–56%). The other six statements were agreed to by approximately two-thirds or over, or by one-third or less of the respondents. Chi-square tests of the School B data are reported in ADI Tables D, E, F, G. The associated C values ranged from .18 to .28, all N s = 524, all p s < .001, tables were 2×3 .

Hypothesis 3

Individuals who deviate from socially prescribed norms will tend to attribute their own deviation to external forces, while individuals who do not deviate will tend to believe that those who do, do so because of internal forces (e.g., a personality with character defects). For example, if *I* cheat, I would claim it is because of unfair grading methods, because others also cheat, etc. (external factors); while if *you* cheat, and I don't, it is because you are "lazy," have a "weak character," etc. (internal factors).

Attitude items blaming cheating on external circumstances tended to be judged less condemning of cheating than did those items that attributed cheating to internal, or character defects. Therefore, if cheaters

agree with items explaining cheating to be a result of internal or personality factors (while noncheaters respond oppositely), this does not unequivocally confirm Hypothesis 3. Cheaters would be expected to prefer moderately condemning statements to more severely condemning statements whether or not the moderate statement blames cheating on the environment.

School A study. The School A test of Hypothesis 3 was open to this criticism. However, the results suggest support (though they do not provide critical evidence) for the hypothesis. The Ss were first divided into two groups on the basis of their tendency to endorse attitude items that blamed cheating on external forces, or to endorse attitude items that placed the blame on the individual (internal forces). Those who scored higher on external than internal items were placed in one group, and those who scored oppositely in a second group. This variable, as contingent upon the cheating score variable, was tested by means of chi square, yielding qualified confirmation: $C = .33$, $N = 165$, $p < .001$, 2×5 table. (See ADI Table H).

School B study. Four categories were distinguished in terms of the Ss' replies to the following two attitude items.

1. "Cheating reveals a basic defect in the character of the student."

2. "Cheating is basically wrong, but is sometimes necessary when too much emphasis is placed upon grades."

Individuals who agreed with the first item and disagreed with the second were designed the "character-defect" group ($N = 169$). Those who responded just oppositely were designated the "social pressures" group (N also equaled 169). Some individuals agreed with both statements (N

$= 82$) and others disagreed with both ($N = 103$).

These last two groups did not enter into the theoretical scheme of the investigation. The only prediction was that the "basic defect" group would have a larger than chance proportion of noncheaters, and that the "social pressures" group would have a larger than chance proportion of active cheaters. Hypothesis 3 was tentatively confirmed: $C = .36$, $N = 338$, $p < .001$, 2×3 table (ADI Table I).

Somewhat greater confidence may be placed in the School B than in the School A results although, again, confirmation is not unequivocal. The two attitude items differ somewhat in the strength of their condemnation of cheating, but both indicate that cheating is "basically wrong" and both items were agreed to by approximately the same percentage of all respondents (53% and 47%, respectively).

Cheating and Sundry Variables

In both schools, the cheater could be typified as being younger than the noncheater, a fraternity member, dependent upon others for financial support, single, a freshman or sophomore, and possessor of low grades. (Associated chi square-based p values were all less than .001.) Whether or not a student claimed church membership ($p > .70$), or was male or female ($p > .20$), was unrelated to cheating behavior.

An honor system for controlling cheating was not popular. This system was preferred by but 28% of the combined samples of students. Most popular with both student bodies was a system in which the faculty do the police work, while the students serve as trial judges (50% of combined samples). Both cheaters and non-

cheaters at School A—the question was not asked at School B—felt that instructor shortcomings (42%) were more responsible for frequent cheating than were factors relating to the nature of the test (23%) or the physical testing environment (35%): chi-square-based $p < .02$.

DISCUSSION

At first blush, the clinical concept of projection would seem sufficient to explain the above-reported findings. Cheaters perceive both themselves and their environment in such a way as to be able to maintain flattering, or at least acceptable ways of regarding themselves.

But individuals also *act upon* their environment and thereby change what is objectively at hand to be perceived. One may thus speak of selective perception *within* an environment and, as well, of a functionally selective choice of environment. And of course a most important way in which environments are selectively chosen is through the friends one chooses and the groups in which one seeks membership.

If a person behaves in a moderately unacceptable way (as, for example, by cheating), either the mechanism of selective perception, or the functionally selective choice of an environment (of friends), may be sufficient to provide that person with adequate protection against unflattering self-thoughts. If a person behaves in a more seriously unacceptable way (as, for example, by indulging in homosexual behavior), it may become necessary for him to “bolster” selective perception (in the sense of projection), with a selective choice of friends who are similarly deviant.

If the selective choice of friends is adequate to the need, one may be able

to “abandon” selective perception of the environment, as it were, and report upon his social world pretty much as it objectively exists—that is, as it would be perceived by an observer who was not ego-involved, and who was in an advantageous position for observation. If this condition existed, a socially deviant person might be able to estimate the extent of deviant behavior in the community at large more accurately than would the nondeviant person.

In reference to the present study, and according to prediction, cheaters more than noncheaters perceived others to frequently cheat. But the tendency for deviants to overestimate the *true* frequency of deviant behavior could not be tested. The data suggest that the deviant individual's estimate may have been *more accurate* than that of the nondeviant. For instance, students at School A who have rarely or never cheated (cheating scores of zero or one), estimated the proportion of “fairly regular” cheaters at their school to be only about 10%; while “frequent cheaters” (those with cheating scores of three or more) estimated the number of their fellow students who cheat “fairly regularly” at around 40%. And yet 81% of the total School A sample admitted having cheated while in college and 46% admitted having cheated during the semester just completed. Although the ambiguity of the phrase “fairly regularly” makes interpretation uncertain, it appears possible that the 40% estimate made by the cheaters may have been more accurate than the 10% estimate of the noncheaters.

Two studies of sexual deviancy—Finger (1947), and Melikian and Prothro (1954)—further illuminate this interpretational difficulty. These

investigators used American and Arab male college students as Ss. In both studies, the Ss were asked about their own sex practices, and they were also asked to estimate the frequency of these practices among friends, and among adult male members of their community. Those Ss who admitted deviant sex practices tended to believe that others also engaged in these same practices, and their estimates were higher than those given by Ss who denied indulging in deviant sex practices.

The results of Finger, and of Melikian and Prothro, appear to provide additional clear cases of projection in the clinical sense of that term. And yet deviant Ss in both studies made more *accurate* estimates of the true frequency of deviant sexual behavior than did nondeviants, as this was gauged against the frequency with which these same Ss admitted having themselves indulged in deviant sex practices. For example, pairs of estimates of frequency of deviant sexual activity among adult males in the community were available from these authors' data. There were two cultures (American and Arab), and three "deviant" acts (masturbation, homosexuality, and heterosexuality outside of marriage), thus providing six bases for comparing the estimates of deviancy made by deviants and by nondeviants. In five of those six cases, deviants gave more accurate estimates than did nondeviants. The exception concerned extramarital heterosexuality in the case of American Ss. Deviants and nondeviants both

overestimated, but deviants did so by a factor of about 1.6 and nondeviants by a factor of about 1.4.

In the present study, both those labeled cheaters and those labeled noncheaters estimated less cheating among friends than among the college population in general (School A study). The median estimates were 23% and 10% respectively. However, the cheater group lowered their estimates less (going from college population to friends), than did the noncheater group, though not significantly. This also would be predicted from "small group theory" more readily than from "perceptual defense theory."

It appears that the results of this study and the studies of Finger and of Melikian and Prothro, can be as well explained by the forces determining group membership, and group attitude and belief systems, as they can by a perceptual defense type of hypothesis.

REFERENCES

- DUNNETTE, M. D. Fads, fashions, and fold-rol in psychology. *American Psychologist*, 1966, **21**, 343-352.
- FINGER, F. W. Sex beliefs and practices among male college students. *Journal of Abnormal and Social Psychology*, 1947, **42**, 57-67.
- FRENKEL-BRUNSWIK, E. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality*. New York: Ronald, 1951. Pp. 356-419.
- MELIKIAN, L., & PROTHRO, E. T. Sexual behavior of university students in the Arab Near East. *Journal of Abnormal and Social Psychology*, 1954, **49**, 59-64.

(Received December 9, 1966)

COOPERATIVE VERSUS COMPETITIVE DISCUSSION METHODS IN TEACHING INTRODUCTORY PSYCHOLOGY

DONALD BRUCE HAINES AND W. J. McKEACHIE¹

University of Michigan

Cooperative and competitive techniques of teaching discussion sections of general psychology were compared with respect to their effects on student anxiety, student achievement, and student satisfaction. The experiment involved 4 sections of introductory psychology. Students in these sections participated in class discussions conducted in a competitive manner for 2 weeks and with a cooperative method for 2 weeks. The competitive condition resulted in higher tension, poorer achievement in recitation, and less satisfaction than the cooperative condition.

How should a college discussion section be conducted? The present research was a direct attempt to compare competitively oriented and cooperatively oriented techniques of discussion. The comparison was made in terms of the relative amounts of tension produced by each technique, the effect of each technique on student performance, and the effects of the techniques on student satisfaction and recall.

A definitional scheme suggested by Deutsch (1949) assumes that in a group each member has certain ends he wishes to attain; that is, goals which have high attraction (or valence). To reach these goals, the individual may or may not have to depend on the behavior of other group members. In cooperation, what

an individual does to help himself helps others; in competition, anything one does to help himself prevents others from moving toward their goals. In Deutsch's experiment, there was no clear difference between competitive and cooperative groups in terms of the amount of individual learning.

Hypotheses

1. There will be a higher level of group tension in competitive conditions than in cooperative conditions.

Both cooperative and competitive groups have tensions increased by the nature of the task. However, there is a difference in the instrumental behaviors leading to tension perceived by the individuals in the competitive group relative to the cooperative groups. Individuals in cooperative groups have many different paths to achieving their goals. This is not true for those in competitive groups. Consequently there is much greater probability that cooperative group members will have tension associated with the task reduced by the action of other members. The reverse is true for competitive groups. Here the probability is high that tension as-

¹ Donald Bruce Haines was killed in a plane crash in Ethiopia in August 1965 while doing research for the Aerospace Medical Laboratories of Wright-Patterson Air Force Base. This abridgement of his doctoral dissertation was prepared by W. J. McKeachie, his doctoral thesis chairman at the University of Michigan, who benefited from suggestions by Theodore Newcomb and members of the staff of the project of the United States Office of Education Research Contract O.E. No. SAE-8451 to W. J. McKeachie, J. E. Milholland, and R. L. Isaacson.

sociated with task will be increased by the activity of other members.

2. The higher tension levels in competitive conditions will result in disruption of performance in those conditions as compared with performance in a cooperative condition.

3. Unreduced tensions associated with tasks begun by one individual and completed by another in the group will be greater for competitive conditions than for cooperative conditions.

Under cooperation, the completion of a task moves everyone toward the goal, regardless of who began the task, but this is not so under competition. There only the person completing the task moves toward his goal.

4. The unpleasantness of high tension levels in competitive conditions relative to tension levels in cooperative conditions will result in member preference for and satisfaction with cooperative over competitive conditions in the classroom.

METHOD

Subjects

Eighty-two undergraduates enrolled in four sections of the introductory psychology course at the University of Michigan were used as subjects. The subjects were told that 4 weeks of the semester would be treated as experimental sessions in the sense that particular grading policies would be followed and that they would be observed during this period by two graduate students sitting in the class.

To check comparability of groups, a number of measures were gathered on the second day of class. Intellectual capacity was assessed by obtaining individual student American Council for Education (ACE) scores from the Bureau of Psychological Services. Class level, area of concentration, age, and sex were taken from a personal data sheet circulated to all students in the experimental groups. Median breaks were made for each of these variables and a chi-square test for differences between sections was conducted. For none of the variables did differences approach significance.

Enrollment figures were as follows: two sections of 20 students each, and two sections of 21 each.

Variables

The independent variable for this research was teaching technique, which was varied so as to result in a competitive atmosphere and a cooperative atmosphere. In both conditions, students were instructed that part of their final grade was dependent upon their recitation performance, but in the cooperative sessions they were also told that anything one individual did to help himself reach the goal (*viz.*, answering recitation questions correctly, thus getting a higher final grade) automatically moved everyone closer to each of their goals (*i.e.*, when one student answered a question correctly, everyone got credit toward his individual final grade). In the competitive situation, however, each student's grade depended on how well he answered questions as compared with the other students.

Control for content and sequence was achieved by adopting the following balanced design:

Classes 1 and 3: 2 weeks of cooperative technique + 2 weeks of competitive technique.

Classes 2 and 4: 2 weeks of competitive technique + 2 weeks of cooperative technique.

Two instructors each taught two classes. Each class was its own control.

Dependent Variables

Dependent variables fall into three categories: (a) assessments of tension level, (b) assessments of performance, and (c) assessments of satisfaction.

Assessments of Tension Level

Three procedures were used to get at the arousal and maintenance of tension systems in the individual and in the group. Two items on a questionnaire assessed tension level in the individual: "This technique promotes an easy, relaxed atmosphere in class." "This technique made me feel anxious and uneasy."

The second means of assessing tension levels in the group consisted of the use of two independent observers who categorized interaction of subjects during the experimental sessions in terms of the Fouriezos, Hutt, and Guetzkow (1950) observational technique.

The final means of assessing the unre-

duced tensions in the group consisted of a measure of the Zeigarnik effect produced when students volunteered for questions, but were unable to answer them completely. The Zeigarnik effect is observable in group functioning. Horwitz (1954) tested this by seeing if the Zeigarnik effect existed for individuals in group goal situations and demonstrated that it did. It would, then, be expected in our study that the tension aroused by attempting an answer and failing to complete it would be more likely to be reduced by the success of another student in the cooperative group than in the competitive group.

A Zeigarnik Recall Form was constructed which simply asked each student to recall as accurately as possible each question he volunteered for and tried to answer, whether credit was given or not. These forms were circulated and collected during the final experimental session of each condition.

Assessment of Performance

A daily measure of recitation performance was obtained by providing each instructor with a seating chart labeled with the names of students in each of his sections, and then having him score students attempting answers to questions. The procedure selected for discussion periods was that of recitation-drill, following the form described by Guetzkow, Kelly, and McKeachie (1954).

The recitation-drill approach consists of bringing to class a prepared list of questions on specifically assigned textbook material. Depending upon the experimental conditions, the student or his group was given credit for answering the question correctly. If the answer was incorrect or incomplete, the credit was lost. The instructor kept asking the question until someone replied correctly. If no one was able to do so, the whole group lost credit in the cooperative condition, and as many students as volunteered lost credit in the competitive condition. When everyone volunteered in both situations, the amount lost or gained was identical. To the extent that not everyone volunteered in the competitive group, then the cooperative group suffered a heavier penalty. If anything, this inequity operated against the major hypothesis, and hence was allowed to remain in the design. The final measure of group performance for all conditions consisted of the number of questions covered per minute for a given session.

The other measure of group performance consisted of two scores on the regular course hour-long examination. The hour examination consisted of two parts: Part 1, covering the first 2 weeks of the course (which for two sections consisted of the cooperative condition and for two was the competitive condition), and Part 2, which covered the last 2 weeks of the experimental sessions. The examination was multiple-choice, containing 40 questions with 20 to each part.

Assessment of Satisfaction

The following items of the student post-session questionnaire comprised the measure of satisfaction: (a) "I preferred this technique to the other one"; (b) "This technique made me doubt my own abilities and lowered my self-assurance"; and (c) "I would enjoy being taught by this technique."

RESULTS

The experimental results support the hypotheses. The results indicate that the independent variable of discussion technique was effectively manipulated, and that crucial control measures did in fact remain constant across sections. In all, 18 predictions were made concerning manipulations, controls, and dependent measures. Of these, 15 were significantly supported, two were in the predicted direction but were not significant, and one of the 18 predictions was neither significant nor in the predicted direction.

Levels of Tension

The major hypothesis for this research stated that relative to cooperative conditions in the classroom, competitive conditions produce an excessive tension level, as evidenced by greater incidence of self-oriented need, lack of relaxed atmosphere, and feelings of anxiety and uneasiness. Students in the competitive sessions consistently showed a greater incidence of self-oriented need per act (as assessed by observers) compared with their behavior in the cooperative

TABLE 1

INCIDENCE OF SELF-ORIENTED NEED: OVER-ALL ASSESSMENT PER SESSION FOR COOPERATIVE VERSUS COMPETITIVE TECHNIQUES

M_D^a	SD_D	Direction of prediction	t^b
2.68	1.12	Comp. Coop.	9.4*

^a Incidence between conditions. (Each class paired with itself—1st session cooperative with 1st session competitive, etc.)

^b t test of direct differences computed. $N = 16$.

* $p < .001$.

sessions. ($F = 84.50$, $df = 1/24$, $p < .001$.)

Similar results hold for the incidence of self-oriented need assessed as an overall measure for each session.

Questionnaire items (postsession questionnaire) relating to anxiety perceived by the individual also clearly bear out the hypothesis. Students in the competitive sessions felt distinctly more tense and anxious than they did in the cooperative sessions. (See Table 2.)

In general then, we have provided

TABLE 2

DIFFERENCES IN SELF-REPORTED ANXIETY AND TENSION DURING DISCUSSION

Questionnaire item	Means		Difference	t
	Com- petitive	Coop- erative		
3. This technique made me doubt my own abilities and lowered my self-assurance	+ .84	-1.01	1.85	7.72*
4. This technique promotes an easy, relaxed atmosphere in class	-2.18	+1.97	4.15	23.48*
6. This technique made me feel anxious and uneasy	+1.26	-1.51	2.77	11.02*

Note.—Likert-type scale. Strongly Agree (+3) to Strongly Disagree (-3).

* $p < .01$.

documentation that the utilization of competitive discussion procedures under the controlled conditions specified in this research can and does lead to high levels of tension in the classroom.

Disruption of Performance

In daily recitation, the performance measure was the number of questions covered during the class session. Students in the cooperative sessions covered more questions than they did in the competitive session. ($F = 5.65$, $df = 1/24$, $p < .05$.) Examination performance of the two groups, however, did not differ significantly.

Unreduced Tensions

The hypothesis that there should be less unresolved tension in cooperative than competitive classes was tentatively supported. ($F = 3.97$, $df = 1/24$, $p < .10$.)

Satisfaction

The final prediction of the present research was that tension levels high enough to lead to anxiety, incidence of self-oriented need, and disruption of performance would result in lowered satisfaction and preference for the conditions resulting in these un-

TABLE 3

SATISFACTION WITH AND PREFERENCE FOR THE DISCUSSION TECHNIQUES

Questionnaire item	Means		Mean difference	t^b
	Com- petitive	Coop- erative		
2. I preferred this technique to the other one	-.82	+1.38	+2.20	5.78*
3. I would enjoy being taught by this method	-1.06	+1.17	+2.23	7.16*

Note.—Likert-type scale. Strongly Agree (+3) to Strongly Disagree (-3).

* t test of direct differences was computed. $N = 82$.

* $p < .01$.

pleasant consequences. Table 3 illustrates that students did prefer the cooperative method.

DISCUSSION

The nature of the goal interdependency structured in the college classroom has a powerful effect upon student behavior. Today's student has so many conflicting demands placed upon him by extracurricular activities and varied social responsibilities that the classroom atmosphere must be potent indeed to attract and hold even a part of the student's interests. The prevailing answer has been to promote deliberately a keenly competitive atmosphere in the classroom with the hope that students will be motivated and devote more energy to school work.

It is presumptuous to question the effectiveness of competition, especially when it is so deeply implicit in modern education, unless a sound basis for demonstrating otherwise can be established. The present research has provided such a basis. Competition, in particular a competitive grading policy, arouses tension in the individual. Many find it agreeable that competition should do so. What is little recognized, however, is that the contribution made by competition

to tensions already existing in the student is so great that undesirable consequences may follow. The present research demonstrated that students in competitive discussion situations became more anxious, displayed a greater incidence of self-oriented needs, and found themselves losing self-assurance. Further, they were less able to perform effectively in recitation, and they became dissatisfied with the discussion procedure. When the discussion was structured cooperatively, students felt less tense, displayed more task-oriented behavior, worked more effectively, and enjoyed the discussion.

REFERENCES

- DEUTSCH, M. A theory of cooperation and competition. *Human Relations*, 1949, **2**, 129-152.
- FOURIEZOS, N., HUTT, M., & GUETZKOW, H. Measurement of self-oriented need in discussion groups. *Journal of Abnormal and Social Psychology*, 1950, **45**, 682-690.
- GUETZKOW, H., KELLY, E., & McKEACHIE, W. An experimental comparison of recitation, discussion and tutorial methods in college teaching. *Journal of Educational Psychology*, 1954, **45**, 193-207.
- HORWITZ, M. The recall of interrupted group tasks: An experimental study of individual motivation in relation to group goals. *Human Relations*, 1954, **7**, 3-38.

(Received December 16, 1966)

Oxford University Press



Fundamentals of Measurement:

TECHNIQUES AND PRACTICES

SECOND EDITION

By N. M. DOWNIE, *Purdue University*

"This text is well-organized and very readable. It goes beyond the usual text by providing a wealth of illustrative material especially with respect to multiple choice and other types of objective test items. It can challenge the very able student and still be meaningful to the bulk of beginning students."—Thomas D. McSweeney, *University of San Francisco*

1967

484 pp.

tables and figures

\$6.50

Teaching the Disadvantaged Child

By SIDNEY W. TIEDT, *San Jose State College*

Directed to those in teacher education, in-service programs for public school teachers, and institutes focusing on the teaching of deprived children, this work provides an instructional guide that suggests effective methods, materials, and media for reaching disadvantaged children and meeting their special needs. The eight chapters, written specifically for this book, are by experts in the various fields of education: John Morlan, Robert Ramonda, Sidney W. Tiedt, Iris M. Tiedt, Weldon R. Parker, Allen C. Friebe, and Charles Carter.

Spring 1968

300 pp.

30 illus.

prob. \$4.50

Special Education:

CHILDREN WITH LEARNING PROBLEMS

By ROGER REGER, WENDY SCHROEDER and KATHIE USCHOLD,
Special Educational Services

This new text is intended for those interested in the formal education of children who deviate from the average in their learning patterns or general behavior as a result of emotional disturbances or minimal brain injuries. Emphasis is on individualized programs with specific outlines for implementing action in the classroom.

Spring 1968

450 pp.

10 illus.

\$6.00

Psychological Diagnosis in Clinical Practice:

WITH APPLICATIONS IN MEDICINE,
LAW, EDUCATION, NURSING, AND SOCIAL WORK

By BENJAMIN POPE and WINFIELD H. SCOTT, *both of The Psychiatric Institute, University of Maryland School of Medicine*

Dealing with the clinical application of psychological tests, this study emphasizes principles of inference and interpretation over operational procedures of assessment. Practitioners of psychology and those in related fields, will find the book useful, as will students during their graduate years in practicum training and in introductory courses in clinical psychology and psychological testing. Extensive references and a selection of illustrations, including ten color plates of the Rorschach inkblot tests, complement the study.

1967

360 pp.

18 illus. (10 in color)

\$8.00

OXFORD UNIVERSITY PRESS

200 Madison Avenue, New York, N.Y. 10016



Harper & Row

PSYCHOLOGICAL FOUNDATIONS OF EDUCATION 2nd Edition

Morris L. Bigge and Maurice P. Hunt

This well-documented text takes a semihistorical and comparative approach. The authors have reworked each chapter to update references, improve continuity and style, and add new material. Part I, unique in scope, treats the biology, psychology, and sociology of human nature. Part II deals with child and youth development, both physiologically and psychologically. Part III, on learning theory, presents a more penetrating discussion of this topic than can be found in competing texts at present. Part IV demonstrates the practical classroom application of theoretical and factual material in the preceding sections of the text. Instructor's Manual. 603 pages. \$9.95. January.

CHILD DEVELOPMENT AND PERSONALITY 2nd Edition

Paul H. Mussen, John Janeway Conger, and Jerome Kagae

There have been over 675 adoptions of the Second Edition of this standard text.

"The first edition by Mussen and Conger was excellent as an undergraduate text. The second edition . . . maintains the standard of excellence. . . . Highly recommended as being literate, well documented, and well conceived."—*Child Development Abstracts and Bibliography* 625 pages. \$8.75.

READINGS IN CHILD DEVELOPMENT AND PERSONALITY

Paul H. Mussen, John Janeway Conger, and Jerome Kagan

These selections follow a chronological and longitudinal plan and deal with critical issues in infancy, the preschool years, the early school years, and adolescence. A brief introductory essay for each section sets the context for the readings and correlates the problems discussed in the sections. Two previously unpublished readings are included. 480 pages. Paper. \$5.50.

PSYCHOLOGY IN THE CLASSROOM, 2nd Edition

Rudolf Dreikurs

This practical manual provides the prospective and in-service teacher with the background information and methods necessary to deal effectively with behavior problems and learning deficiencies of students. Grounded in the philosophy of democracy and the socio-teleological approach of Adlerian psychology, this new edition enlarges on the significance and techniques of group approaches in the classroom, particularly the use of group discussions. 286 pages. Paper. \$3.75. Just Published.

STIMULUS AND RESPONSE

John A. Barlow

Teaching a theory of why and how learning occurs, this entirely original formulation of operant psychology consists of 13 lessons and four experimental projects that can conveniently be carried out by the student without equipment or supervision. The lessons are prepared in a self-instructional style called "conversational programming." Instructor's Manual, with additional evaluative data. 199 pages. Paper. \$4.75. January.

LEARNING AND HUMAN ABILITIES: EDUCATIONAL PSYCHOLOGY, 2nd Edition

Herbert J. Klausmeier and William Goodwin

Almost completely rewritten and thoroughly updated, this edition continues to emphasize the concept of emerging human abilities, thus integrating the treatment of growth and learning. Course outlines have been prepared for the instructor, and suggest a variety of combinations of all or parts of each of the 18 chapters. A Student Evaluation Guide is available to teachers. 720 pages. \$8.95.

A Student Workbook and Adjunct Program, by Herbert J. Klausmeier, William Goodwin, and Robert Conry is available for use with the text. \$3.25.

RECORDING AND ANALYZING CHILD BEHAVIOR

with Ecological Data from an American Town

Herbert F. Wright

This text presents a method of describing the situations and behavior of individuals in their natural habitats. Conceptual foundations of the method as well as results of its application are given. 291 pages. Paper. \$4.50.

Harper & Row, Publishers • 49 E. 33d Street, N.Y.

PROCEEDINGS

of the 75th Annual Convention of the American Psychological Association 1967

The purpose of the **PROCEEDINGS** is to facilitate the dissemination of information regarding current work in major fields of psychology by making available research reports with a minimum of lag and well in advance of the Convention. The 1967 **PROCEEDINGS**, published in July, contains approximately 200 reports distributed across the major subject areas of psychology.

The 1967 Convention **PROCEEDINGS** will contain the contributed papers of the following Divisions:

- | | |
|---|--|
| 1. Division of General Psychology | 14. Division of Industrial Psychology |
| 3. Division of Experimental Psychology | 15. Division of Educational Psychology |
| 6. Division of Physiological and Comparative Psychology | 16. Division of School Psychologists |
| 7. Division of Developmental Psychology | 17. Division of Counseling Psychology |
| 12. Division of Clinical Psychology | 19. Division of Military Psychology |

PRICE: \$9.00 per copy (\$4.50 to APA members)

Copies of the 1966 **PROCEEDINGS** are still available at \$6.00 per copy
(\$3.00 to APA members)

AMERICAN PSYCHOLOGICAL ASSOCIATION

1200 17th Street, N.W.

Washington, D.C. 20036

PLEASE SEND ME _____ copy(ies) of the 1967 **PROCEEDINGS**

☐ Remittance Enclosed ☐ Bill Me (when issue mailed)

NAME _____

ADDRESS _____

Zip Code _____

The 1966 **PROCEEDINGS** was the first issue as a permanent publication. The 1965 issue was part of a research project on Scientific Information Exchange in Psychology.

MNEMONIC SYSTEMS IN RECALL¹

GORDON WOOD

Michigan State University

A series of 5 experiments was conducted to determine if the use of a mnemonic system influences recall. In Experiment I an attempt was made to ascertain what elements of the mnemonic system, if any, were responsible for improving recall. Presentation rate (Experiment II), transfer paradigm (Experiment III), type of list (Experiment IV), and list abstractness (Experiment V) were manipulated to assess whether these variables differentially affected, relative to a control, the recall of Ss utilizing a particular mnemonic system. The major findings were: When Ss were presented with an additional list (peg list) for the learning and recall trial and instructed to make associations, during the learning trial, between the peg words and the words to be recalled (response words), recall of the response words was markedly facilitated. The Ss presented only the response list and instructed to link successive words of the response list during the learning trial also had facilitated recall. An interaction between presentation rate and instruction condition was obtained, suggesting that a slower rate of presentation is optimal for Ss employing a mnemonic system. Instruction condition did not interact with transfer paradigm, type of list, and list abstractness. Implications of these findings were discussed.

The importance of mnemonic devices for the acquisition of associations between verbal units has been suggested by some writers (e.g., Bugelski, 1962; Runquist & Farley, 1964), but the experimental investigation of mnemonic devices has been, for the most part, neglected. The present series of experiments is designed to investigate some of the aspects of memory-training systems (mnemonic systems). The study attempts in part to determine under what circumstances, if any, memory systems facilitate learning, the elements of the mnemonic device responsible for the "facilitated" learning, and the relationships between mnemonic devices and some other

variables which have been demonstrated to affect learning.

Although there are slight variations among mnemonic systems (e.g., Furst, 1957; Nutt, 1941; Roth, 1961), the basic components of the systems are essentially the same. The first phase of most systems involves the memorizing of a series of "pegs." Generally, words are used as pegs and each word is numbered, with associations between numbers and words, for example, one is a *BUN*, two is a *SHOE*, three is a *TREE*, four is a *DOOR*, etc. Following the memorization of the peg words, new words can be memorized by using bizarre imagery to connect the new words (response words) to the peg words. For example, given that *automobile* is the first response word (word to be memorized), the task is to conjure up a bizarre image connecting the first peg, *BUN*, and *automobile*. A possible image might be a 5-foot bun driving an automobile. The same principle is used for all the remaining response words. Once a bizarre image has been

¹Based on a doctoral dissertation submitted to the Department of Psychology, Northwestern University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Special thanks are due Carl P. Duncan, chairman of the author's committee, for his helpful comments and criticisms. Thanks are also due Albert Erlebacher, Winfred F. Hill, and Benton J. Underwood.

formed between a peg word and a response word, rehearsing either the words or the image is believed to be unnecessary. To recall any of the response words, it is supposedly only necessary to recall the well-memorized peg words. The peg words are presumed to elicit readily the visual images previously formed and thus make the response words easily available. Thus, the peg word is considered to be an effective stimulus for the word to be recalled.

There is some evidence to support the claim made by the proponents of mnemonic systems (e.g., Roth, 1961) that the use of a mnemonic system results in superior recall. Wallace, Turner, and Perkins (1957) presented pairs of English words to subjects (Ss) and had them, working at their own pace, form a visual image between the two words. Only one trial was given on each pair. The Ss started with lists of 25 pairs and worked up to lists of 700 pairs. Recall of the response word upon being given the stimulus word was about 99% up to 500 pairs; at 700 pairs recall dropped to about 95%. Although the authors did not utilize a control, this level of recall is greater than what might be expected from a group not employing imagery. Also, there was some evidence to indicate that the time required to make each association decreased with practice. That is, an S given eight lists of 25 word pairs and instructed to form associations as quickly as possible had an average association time per pair of 20.4 seconds on List 1 and an average association time per pair of 3.0 seconds on Lists 6, 7, and 8.

Smith and Noble (1965) tested Furst's (1957) mnemonic system at three levels of meaningfulness. The experimental groups were given a 1-hour lecture-demonstration on Furst's

hook method of mnemonics, followed by 4 days of "private practice" in imagery. All Ss were given 20 serial learning trials on a consonant-vowel-consonant (CVC) list of 10 items. After 24 hours Ss returned for 10 additional trials on the previously presented list. Over all, the experimental groups (those receiving the special training) showed less loss over the 24-hour rest period than the controls. Yet a complex relationship was found between the levels of meaningfulness and the effectiveness of the method. There were no differences between experimental and control groups for highly meaningful lists, large differences for the lists of medium meaningfulness, and small differences for the low meaningful lists.

Balaban (1910) had Ss learn serial lists of 40 words under two different sets of instructions. One-half the Ss were to use a "mechanical" method, while the other half were to form "conscious" connections, by the use of mnemonics, between successive pairs of words. The author found that Ss using mnemonics did better than those instructed to take a "mechanical" approach.

Although the above studies are somewhat supportive of the position that the use of a mnemonic system can facilitate recall, more evidence is clearly needed. It has not been established what elements of the mnemonic system, if any, are responsible for facilitated recall or under what conditions these elements facilitate recall. Also, no attempt has been made to determine what variables influence the performance of Ss employing a mnemonic system. That is, it is conceivable that variables known to influence the performance of Ss uninstructed in the use of a mnemonic system may not affect the perform-

ance of Ss utilizing a mnemonic system.

EXPERIMENT I

The purpose of this experiment was to determine if a memory system employing peg words, imagery which is bizarre, and nonrehearsal of prior images facilitates performance relative to the appropriate control. In addition, the importance of the individual components of the mnemonic system (namely, presence of peg, image, bizarreness of image, and nonrehearsal of previous images) was evaluated. It was not expected that all the components of the mnemonic system would be of equal importance. Since the use of pegs and imagery are the basis of the system, and nonrehearsal of previous images and bizarreness of the image are subsidiary components, the former were expected to produce greater facilitation of performance than the latter.

Method

Design. In four of the seven conditions Ss were given answer sheets on which the peg words were written. The four groups given the peg words differed with respect to the strategy they were asked to use to learn the words that were read to them (response words). Group 1 was told to make a bizarre image which incorporated the response word and peg word, and not to rehearse previously formed bizarre images. Group 2 was given the same instructions as Group 1 except that they were told to rehearse earlier formed images during any time not devoted to forming an image to link a peg and response word. Group 3 was given the same instructions as Group 1 except that they were told to form common (rather than bizarre) images. Group 4 was told to make a verbal association between the word to be memorized and the peg word. For example, the verbal association might involve using a third word as a mediator between the response word and the peg word.

The remaining three groups were not given peg words. Group 5 was instructed to make a bizarre image for each response word. For example, if the first word is *automobile*,

S might form an image of an *automobile* 15 feet high, 2 feet wide, and 4 feet long. Group 6 was told to link response words with a bizarre image. For example, if the first, second, and third response words are *automobile*, *saltshaker*, and *dog* respectively, S might form an image of a large *saltshaker* driving an *automobile* shaped like a *dog*. Group 7 was left to its own devices. That is, these Ss received standard free-recall learning instructions.

To determine if there was a differential practice effect (nonspecific transfer) as a function of the method used to learn the words, each S was presented three different lists. Murdock (1960) has evidence to indicate that the performance of Ss given standard free-recall learning instructions and only a few trials on each successive list does not improve with practice. Thus, one might expect that there would be greater positive nonspecific transfer in the imagery conditions than in the standard free-recall-instructions condition if Ss decrease the time needed to form an image as they become practiced (Wallace et al., 1957). To measure the practice effects, the order of presentation of three lists was completely counterbalanced for each group. Since there were 18 Ss in each condition, there were three replications of each order of list presentation.

Subjects. The Ss were 126 students from introductory psychology classes. Eighteen Ss were assigned to each of the seven conditions by randomizing the order of the seven conditions 18 times, then assigning Ss in the order of their appearance to the randomized list of conditions. In actuality, the two conditions of Experiment III (see below) were run at the same time as the seven conditions of Experiment I. Thus the order of nine conditions was randomized 18 times and 162 Ss were utilized. However, clarity is more easily maintained if only the seven conditions of Experiment I are considered at this time.

Materials. Six lists of 40 words each were used in this experiment (Table A1, Appendix A). All but 21 of the 240 words were taken from the Underwood and Richardson (1956) norms. The additional 21 words were selected with the only restriction that they be, as are the 219 Underwood and Richardson words, relatively high-frequency concrete nouns. These words were selected because concrete words are believed to elicit images more readily than abstract words (Paivio, 1965). The 240 words were randomly divided into six lists of 40 words each with the only restriction that each list con-

tain approximately the same number of words starting with any specified letter. Of the six lists, three were randomly selected to serve as peg lists and three as response lists. The order of presentation of each peg list and response list, and thus the pairings of the peg and response words, was random with the restriction that no peg and response word pair had the same initial letter.

Procedure. The *Ss* were run in groups of seven, all seven conditions being represented in any one group session. On entering the experimental room, *Ss* were seated and then given a five-page handout. On page 1 (cover page) a note cautioned *S* not to turn the cover page until told to do so. On page 2 the special instructions for each condition were printed. That is, each *S* was told how to attempt to learn the lists of response words. The special instructions were in considerable detail; examples of each strategy were given in the instructions. To demonstrate the nature of the instructions, the general instructions given to all groups and the special instructions for Groups 1, 4, 6, and 7 are presented in Appendix B. Pages 3-5 were answer sheets. A peg-word list was printed on each of the answer sheets for *Ss* in Groups 1-4. In any one set of seven *Ss* run at one time, the order of the answer sheets was the same, but the order of the answer sheets, and therefore the order of presentation of the response-word lists, varied from set to set. Each *S* served in only one session. The peg word for any given response word was the same for all *Ss* in Groups 1-4. The three answer sheets for each *S* not receiving peg words were identical.

Prior to the turning of the cover page, *Ss* were given general information about the experiment. They were told the number of lists to be learned (three), number of words per list (40), presentation rate of the list (5 seconds per word), and mode of list presentation (tape recorder). After this information was given, *Ss* were given an opportunity to ask questions. After all questions had been answered, the experimenter (*E*) explained that each *S* was to employ a particular strategy in the learning of the three lists. The *E* stressed the importance of employing only the strategy explained on page 2 of the handouts. The *Ss* were then instructed to turn to page 2 of their handouts and study their instructions. The *E* cautioned the *Ss* not to ask questions about their individual instructions, but, instead, to study their instructions until they fully understood the strategy they were to use. After each *S* informed *E* that the instructions were com-

pletely understood, *E* turned on the Wollensak tape recorder on which the response words had previously been recorded. Following the reading of each list, *E* stopped the tape recorder and asked *Ss* to write as many of the words as they could on their answer sheet. After *Ss* indicated that no more time was necessary, *E* instructed *Ss* to turn to their next answer sheet and *E* started the tape of the next list. Following recall of the third list, *Ss* were cautioned not to discuss the experiment with anyone and then were dismissed.

Results

Each *S's* answer sheets were scored to determine the number of response words correctly recalled, disregarding order of recall. Spelling and number (i.e., singular vs. plural) errors were disregarded in scoring. Homonyms (e.g., steak for stake) were accepted as correct, but incomplete responses (e.g., snake for rattlesnake) and mistakes apparently due to misunderstanding the intended tape word (e.g., teach for peach) were not accepted as correct. These scoring criteria were used for all experiments.

The mean number of words recalled as a function of stage of practice for each of the seven conditions is presented in Figure 1. An examination of Figure 1 reveals that the groups given peg words (Conditions 1-4) performed better than the three groups not given peg words. The group instructed to rehearse previously formed images (Group 2), the group instructed to form common images (Group 3), and the group told to use verbal mediation rather than imagery (Group 4) all performed slightly better than Group 1. Yet, because these differences were not significant by either a Dunnett or Duncan ($p > .05$) range test, using the total number of words recalled for the three lists, this experiment did not support the notion that rehearsing previously formed images, bizarreness of the image formed, or imagery (as

opposed to verbal mediation) significantly influence performance. Moreover, since Groups 1-4 were superior to Group 7 (control) by either a Dunnett ($p < .01$) or Duncan ($p < .001$) range test, it seems plausible to assume that the presence of the peg words was responsible for the facilitated performance.

The performance of Group 6 (instructed to link words with bizarre images) demonstrated that factors other than the presence of peg words can result in facilitated recall. Since Group 6 was not given peg words but still performed at a level significantly (Duncan test, $p < .01$) superior to the control (Group 7), it seems apparent that the presence of peg words is not the sole means by which improved recall can be obtained. Yet the results suggest that the presence of peg words produces a greater facilitation of performance than does giving instructions to link successive response words. Both a Duncan test and a Dunnett test ($p > .05$) on the total number of words correctly recalled indicated that Groups 1 and 6 did not differ significantly, but Groups 2-4 were significantly superior to Group 6 by Duncan's test ($p < .05$).

As can be seen in Figure 1, the performance of Groups 5 and 7 was markedly inferior to that of the other five groups. The difference between Groups 5 and 7 was not significant, but both of these groups were significantly inferior (Duncan, $p < .001$) to the other five groups. Since Groups 5 and 6 differed only with respect to the instruction to link words with a bizarre image (Group 6) or to form a distinct bizarre image for each word (Group 5), it is probable that linking of items during learning is responsible for the facilitated recall of Group 6.

Since the groups given peg words were told to link the peg words and response words during learning, the peg and response words should be correctly paired at recall. The mean number of words recalled in the correct position (i.e., correctly paired with the appropriate peg word) was computed for Groups 1-4. This measure strongly supported the notion that Ss linked the peg words and response words during learning in that the mean number of misplaced items per list was only .54, .43, .26, and .24 for Groups 1-4 respectively. These groups not only recalled more words, but they recalled them in the correct order. The mean number of words recalled in the correct position (i.e., in the correct blank of the answer sheet) over all three lists was only 3.44, 6.28, and 4.67 for Ss in Groups 5-7. Thus, if the mean number of words recalled in the correct position is used as a response measure, Groups 1-4 were vastly superior to Groups 5-7.

The number of words correctly recalled as a function of serial position was obtained to determine whether instruction condition influenced the customary bowed serial-position effect of free recall (Deese & Kaufman, 1957). The serial-position curves for the first list presented to S are given

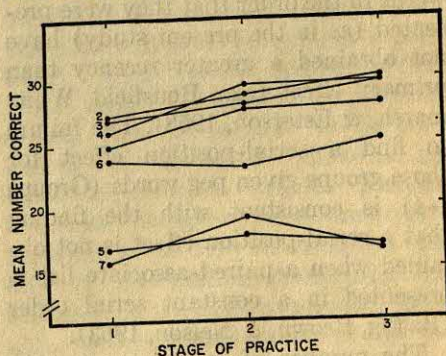


FIG. 1. Mean recall as a function of stage of practice for the seven instruction conditions of Experiment I.

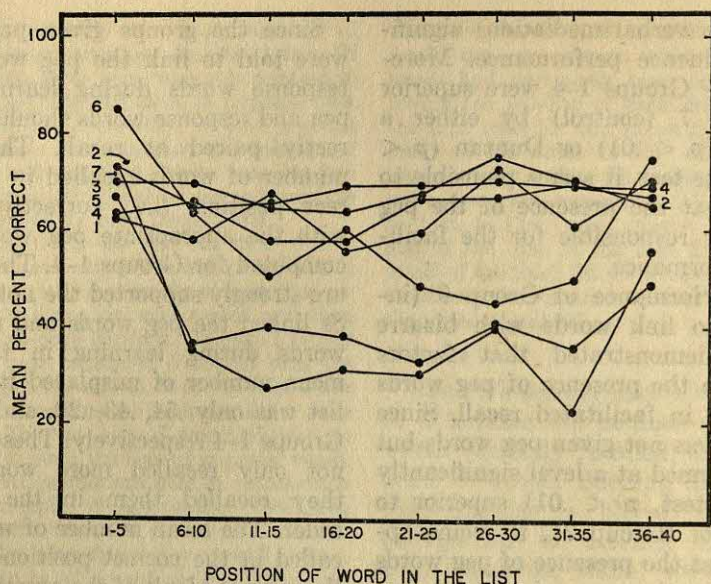


Fig. 2. The serial position curves for the seven instruction conditions of Experiment I.

in Figure 2. From this figure it can be seen that the groups not given peg words (Groups 5-7) showed the typical bowed serial-position effect while the groups given peg words (Groups 1-4) did not. Groups 5-7 had as large a primacy as a recency effect. Although other studies (e.g., Murdock, 1962) have generally found a greater recency than primacy effect, Murdock points out that the studies in which Ss were instructed to recall the words in the order that they were presented (as in the present study) have not obtained a greater recency than primacy effect (e.g., Bousfield, Whitmarsh, & Esterson, 1958). The failure to find a serial-position effect for those groups given peg words (Groups 1-4) is consistent with the finding that a serial-position effect is not obtained when a paired-associate list is presented in a constant serial order (Battig, Brown, & Nelson, 1963).

The number of intrusions per list and the intrusion rate per list were computed to determine if the condi-

tions differed with respect to the eliciting of erroneous responses. The mean number of intrusions per list was 2.93, 3.00, 2.91, 2.83, 1.65, 1.33, and 2.04 for Groups 1-7 respectively. A Duncan's range test on the total number of intrusions revealed that the groups given peg words (Group 1-4) had significantly ($p < .05$) more intrusions than the two groups told to use bizarre imagery but not given peg words (Groups 5 and 6). No other differences were significant. However, the mean number of intrusions may not be the most appropriate measure of overt errors because Groups 1-4 had greater opportunity for intrusions in that they had higher mean recall than the other groups. To equate for total recall, an intrusion-rate measure was obtained by dividing the number of intrusions by the total number of responses made (both correct and incorrect) for each S. The mean intrusion rate was .105, .097, .096, .092, .090, .056, and .080 for Groups 1-7 respectively. A Duncan's range test

indicated that Group 6 had a lower intrusion rate than Groups 1-3 ($p < .05$); all other differences were not significant. Thus, although the intrusion and intrusion-rate measures are not in complete agreement, these measures suggest that Ss instructed to link successive response words during the learning trial have fewer intrusions and a lower intrusion rate than Ss provided with and instructed in the use of peg words.

An analysis of variance (Latin-square design) was made on the number of correct responses to determine whether there was a differential practice effect as a function of the strategy Ss were instructed to employ to learn the response words. The summary table of this analysis is presented in Table 1. The analysis of the ordinal position effect revealed a significant main effect and a Groups $\times$ Ordinal Position interaction. Since an examination of Figure 1 suggested that those groups given peg words showed an improvement with practice, while those not given peg words did not, an orthogonal comparison was made to determine if the significant Groups $\times$ Ordinal Position interaction was due largely to greater practice effects for groups given peg words than for those groups not given peg words. That is, Groups 1-4 were compared with Groups 5-7 (Comparison A, Table 1) to test the notion that the presence of the peg words might account for much of the differences in practice effects. As can be seen in Table 1, this comparison accounts for a considerable amount of the Groups $\times$ Ordinal Position interaction. Moreover, the comparison is significant for both the linear and quadratic components of the trend.

It should be noted that an assumption of the Latin-square design may have been violated in the present

TABLE 1
ANALYSIS OF EXPERIMENT I

Source of variance	df	MS	F
Groups (G)	6	1441.21	18.25***
Sequence (S)	5	14.88	—
G $\times$ S	30	147.12	1.86*
Subjects/G,S	84	78.95	
Lists (L)	2	12.76	1.68
Ordinal position (OP)	(2)	(162.20)	(21.40)***
Linear	1	262.10	34.58***
Quadratic	1	62.29	8.22**
G $\times$ OP	(12)	(19.38)	(2.56)**
Between G linear	(6)	(17.62)	(2.32)*
Comp. A	1	43.89	5.79*
Residual	5	12.36	1.63
Between G quadratic	(6)	(21.14)	(2.79)*
Comp. A	1	33.22	4.38*
Residual	5	18.72	2.47*
G $\times$ L	12	11.51	1.51
Square uniqueness	8	25.27	3.33***
G $\times$ Square uniqueness	48	12.58	1.66*
Pooled residual	168	7.58	
Total	377		

* $p < .05$.
 ** $p < .01$.
 *** $p < .001$.

experiment. The model for the analysis in Table 1 assumes that there is no interaction between ordinal position and lists. If this interaction, in fact, exists, then the ordinal position effect might be confounded. A partial test for this interaction is to test square uniqueness with the pooled residual. Since square uniqueness is significant, there is some reason to question the assumption of no interaction between ordinal position and lists. However, there is also reason not to be too concerned in that all possible sequences (complete counterbalancing of lists and ordinal position) were used in the present experiment. Grant (1948) maintains that when there is complete counterbalancing there is no need to be concerned with the assumption of no interaction between ordinal position and lists. However, since subsequent experiments of this series afford additional information concerning the influence of the ordinal position variable, further consideration of this issue may be unnecessary.

In summary, the results of Experiment I indicated that, with the pres-

ent materials, providing Ss with peg words or instructing them to link successive response words during presentation resulted in superior recall relative to the appropriate control. The use of peg words enabled Ss to recall words in the same order that they were presented; correct ordering of recall was not obtained when peg words were not provided. There was some indication that instructing Ss to link successive response words resulted in fewer intrusions and a lower intrusion rate than providing them with and instructing them in the use of peg words. The serial-position curves obtained were generally in agreement with the results of other investigators. Although there was some question whether a necessary assumption of the underlying statistical model had been satisfied, the results suggested that groups given peg words improved with practice, but groups not given peg words did not.

EXPERIMENT II

The purpose of this experiment was to investigate how the performance of Ss employing a mnemonic system is influenced by presentation time per item. It was believed that Ss required to form images or utilize mediators may benefit more from a longer presentation time per item than Ss not required to form images or use mediators. That is, if the use of mnemonics results in "better learning" and if using mnemonics (forming images or utilizing mediators) requires more time than not using mnemonics, then Ss utilizing a mnemonic system should be able to benefit more from a slow presentation rate than Ss not employing a mnemonic system. Although there is little evidence to indicate what the optimal presentation rate is (if, in fact, there is one) for Ss using

mnemonics, evidence from the Wallace et al. (1957) study suggests that the optimal rate is slower than 3 seconds. That is, in the Wallace et al. study a practiced S required an average of 3 seconds to make a "mnemonic" association between two words. Thus, it should follow that a presentation rate of over 3 seconds is likely to be more nearly optimal than a presentation rate of less than 3 seconds for relatively unpracticed Ss.

Method

Design. Two conditions were run in this experiment: a group given the peg words and complete imagery instructions (i.e., like Group 1 of Experiment I), and a group given standard recall instructions (i.e., like Group 7 of Experiment I). Both groups were presented the words at a rate of 2 seconds per word. Comparisons were made between these two groups and the corresponding groups of Experiment I. The design is a 2×2 factorial with instructions (mnemonic system) as one variable and the presentation time per word as the second variable. The prediction is that the interaction between these two variables will be significant such that there will be a greater difference in performance between the mnemonic system condition and the nonmnemonic system condition at the 5-second presentation rate than at the 2-second rate.

Procedure and materials. The procedure and materials were the same as they were for the corresponding groups of Experiment I, except that the response words were presented at a 2-second rather than a 5-second rate. Since only two conditions were used, Ss were run in groups of six, three Ss in each condition. The six Ss receiving the same order of the lists were run at the same time. Experiment II was run concurrently with Experiment I. That is, Ss were randomly assigned to serve in Experiment I or Experiment II solely as a function of their order of appearance in the laboratory.

Results

The mean number of words recalled as a function of stage of practice for Conditions 1 and 7 of Experiments I and II is presented in Figure 3. As expected, presentation rate (2 seconds

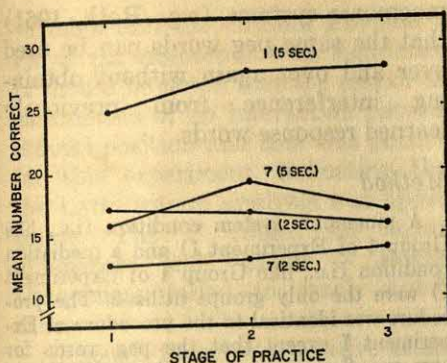


FIG. 3. Mean recall as a function of stage of practice for the four groups of Experiment II. (Presentation rate—2 versus 5 seconds—and instruction condition—1 versus 7—were the variables manipulated.)

versus 5 seconds) and instructional condition (Condition 1 versus 7) both significantly influenced performance (Table 2). The most important result of this experiment, however, is the significant interaction between Presentation Rate $\times$ Instruction Condition. A graphical presentation of this interaction is represented in Figure 4 by plotting the mean recall over all lists as a function of presentation rate. Thus, this study indicates that

the relative effectiveness of a mnemonic system strategy depends in part on the rate of presentation of the items.

The failure to find an interaction between instruction condition and ordinal position in this experiment (see Table 2) and the presence of this interaction in Experiment I suggests that presentation rate may be a factor determining whether Ss utilizing a mnemonic system improve with practice. When a 5-second presentation rate was used, Ss in Condition 1 improved with practice; however, Ss in

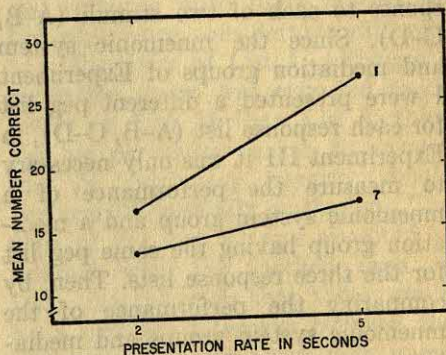


FIG. 4. Mean recall over all three lists for Instruction Conditions 1 and 7 at two presentation rates.

TABLE 2
ANALYSIS OF EXPERIMENT II

Source of variance	df	MS	F
Presentation rate (PR)	1	2787.85	43.75***
Instructions (I)	1	2373.41	37.25***
Sequence (S)	5	44.83	—
PR $\times$ I	1	541.50	8.50**
PR $\times$ S	5	55.24	—
I $\times$ S	5	226.95	3.56**
PR $\times$ I $\times$ S	5	122.47	1.92
Subjects/PR,I,S	48	63.72	—
Ordinal position (OP)	2	126.46	18.12***
Lists (L)	2	9.86	1.41
PR $\times$ OP	2	6.80	—
I $\times$ OP	2	6.36	—
PR $\times$ L	2	9.34	1.34
I $\times$ L	2	17.50	2.51
I $\times$ PR $\times$ OP	2	13.55	1.94
I $\times$ PR $\times$ L	2	6.12	—
Square uniqueness	8	8.63	1.24
I $\times$ Square uniqueness	24	17.34	2.48**
Pooled residual	96	6.98	—
Total	215		

* $p < .05$.
 ** $p < .01$.
 *** $p < .001$.

Condition 1 did not show an improvement with practice when a 2-second presentation rate was used. Condition 1 of Experiment I and Experiment II were analyzed to determine if the apparent improvement with practice depended on presentation rate. Since this interaction was not significant ($F = 1.18$, $df = 2/48$, $p > .05$), the notion that improvement with practice for Ss in Condition 1 depended on a slow rate of presentation was not supported.

EXPERIMENT III

The purpose of this experiment was to compare a mnemonic system group

(namely, instructed to use a bizarre image to connect each peg and response word, and cautioned not to rehearse previously formed images) and a mediation group (namely, instructed to find a verbal means, such as using a third word as a mediator, to connect each peg and response word), in a negative transfer paradigm. Numerous studies (e.g., Twedt & Underwood, 1959; Underwood, 1951) have demonstrated that *Ss* learning two different successive responses to the same stimulus (A-B, A-D) have a more difficult time with the second pair than *Ss* learning a different response to each of two stimuli (A-B, C-D). Since the mnemonic system and mediation groups of Experiment I were presented a different peg list for each response list (A-B, C-D), in Experiment III it was only necessary to measure the performance of a mnemonic system group and a mediation group having the same peg list for the three response lists. Then, by comparing the performance of the mnemonic system groups and mediations groups of Experiment I and Experiment III it is possible to determine whether a mnemonic system strategy or mediation strategy results in the greatest negative transfer. That is, it should be possible to determine whether the effect of having the same pegs for consecutive lists of response words produces a greater negative effect when *S* uses a bizarre image to connect each peg and response word or when *S* uses some form of verbal mediation.

The interfering effects of the first list are expected to be greater in the mediation condition, thus resulting in greater negative transfer for the mediation condition. The reasoning for the prediction, although there is no evidence to support it, is based on the claim made by the advocates of

mnemonic systems (e.g., Roth, 1961) that the same peg words can be used over and over again without obtaining interference from previously learned response words.

Method

A mnemonic system condition (i.e., like Group 1 of Experiment I) and a mediation condition (i.e., like Group 4 of Experiment I) were the only groups utilized. The procedure was identical to the procedure of Experiment I except that the peg words for any one *S* were the same for each of the three different response lists. One-third of the *Ss* ($N = 6$) in each condition had Peg List 1, one-third Peg List 2, and one-third Peg List 3. As was pointed out earlier, these two conditions were run at the same time as the seven conditions of Experiment I.

Results

The mean number of words correctly recalled for the four groups as a function of stage of practice is presented in Figure 5. This figure clearly indicates that the performance of *Ss* given the same peg words for successive lists (i.e., A-B, A-D) decreased with practice while the performance of *Ss* given different peg words for successive lists (i.e., A-B, C-D) increased with practice. These differences were significant (Table 3), that is, the ordinal position (practice)

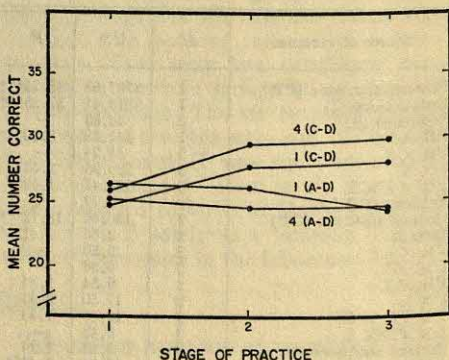


FIG. 5. Mean recall as a function of transfer paradigm (A-B, C-D versus A-B, A-D) and stage of practice for Instruction Conditions 1 and 4 of Experiment III.

variable and the Ordinal Position $\times$ Paradigm interaction were significant. It should be noted that the test of the assumption of no interaction between ordinal position and lists was positive for this experiment, indicating that the Latin-square analysis was appropriate. The F values for square uniqueness and Groups $\times$ Square Uniqueness failed to reach significance.

Although there was positive transfer with the A-B, C-D paradigm and not with the A-B, A-D paradigm, there was no evidence of differential transfer as a function of instruction condition (Table 3). The prediction that Ss using imagery would be able to use the same peg words to learn successive lists without obtaining interference effects was not supported. In addition, there was no evidence that the absolute recall level differed for the two instruction conditions. An analysis of the number of words recalled in the correct position, the number of intrusions and the intrusion rate also failed to reveal any differences between the bizarre imagery and mediation condition.

TABLE 3
ANALYSIS OF EXPERIMENT III

Source of variance	df	MS	F
Paradigm (P)	1	352.67	3.21
Instructions (I)	1	5.35	—
Sequence (S)	1	254.33	2.31
P $\times$ I	1	75.85	—
P $\times$ S	5	26.60	—
I $\times$ S	5	32.09	—
P $\times$ I $\times$ S	5	102.25	—
Subjects/P,I,S	48	100.88	—
Ordinal position (OP)	2	36.81	3.15*
Lists (L)	2	2.24	—
P $\times$ OP	2	123.18	10.53***
I $\times$ OP	2	5.84	—
P $\times$ L	2	42.89	3.67*
I $\times$ L	2	2.02	—
I $\times$ P $\times$ OP	2	2.67	—
I $\times$ P $\times$ L	2	4.96	—
Square uniqueness	8	18.74	1.60
I $\times$ Square uniqueness	24	11.70	1.00
Pooled residual	96	11.70	—
Total	215		

* $p < .05$.
*** $p < .001$.

EXPERIMENT IV

This experiment was conducted to determine if instruction condition would interact with the type of list used. An interfering (I) and a mediation (M) list were constructed by using normative data. Because of the nature of the words selected, Ss employing a mediation approach (i.e., instructed to verbally connect each peg and response word) were expected to do well on List M and poorly on List I. That is, it was expected that Ss would have little difficulty in thinking of words to mediate the peg and response word pairs in List M, but experience considerable difficulty with the pairs of List I. Although a mediation group was expected to do better on List M than List I, there was no particular reason to suspect that Ss employing imagery should perform differentially on the lists. The type of list learned was expected to interact with the method used to learn the list such that there would be a marked difference between the performance on the two lists under one method (verbal mediation) and not under the other (bizarre imagery).

Method

All 36 Ss in each of the two conditions, bizarre imagery and verbal mediation, learned the two lists using a procedure essentially identical to the procedure used in Experiments I-III. The Ss were given a four-page handout consisting of a cover page, special instructions regarding strategy, and two answer sheets. A list of peg words was printed on each answer sheet. Each response-word list was presented on a tape recorder at a rate of 5 seconds per word. The order of presentation of the response lists was counterbalanced.

List I was constructed by using 40 stimulus words (e.g., TABLE) and the dominant response to each stimulus word (e.g., chair) from the Minnesota norms (Russell & Jenkins, 1954). The stimulus words were used as peg words and the dominant responses were used as response words. However, the

peg words and response words were repaired. For example, the response word to the peg word TABLE was *crackers* instead of *chair*. The learning of this list was expected to be difficult for Ss using a mediational approach because the dominant response to each peg word may be an incorrect response. If TABLE elicits *chair*, *chair* might interfere with the learning of the correct response word, *crackers*.

List M was constructed by using 80 words (40 pairs) from the Underwood and Richardson sensory impression norms. Pairs of words were selected which elicit the same sensory impression response (e.g., CIGAR and ETHER both elicit *smelly* as a dominant sense impression). The Ss using a mediational approach were expected to do well on List M because previous pilot work using a similar list indicated that Ss are able to supply appropriate mediators. For example, Ss asked to learn the response *cigar* to the stimulus ETHER reported using the mediator *smelly*. Lists I and M are presented in Table A2, Appendix A.

Results

The mean number of words recalled for Condition 1 (bizarre imagery) was 27.22 and 33.00 for Lists I and M respectively. For Condition 4 (verbal mediation) mean recall was 25.17 and 30.89 for Lists I and M respectively. List M was significantly easier than List I ($F = 56.84$, $df = 1/32$, $p < .01$). The only other significant effect was that of ordinal position ($F = 10.04$, $df = 1/32$, $p < .01$). Since there was an increase in performance as a result of practice, this experiment is consistent with the earlier three experiments in demonstrating a practice effect for those groups given peg words and a 5-second presentation rate. The predicted interaction between instruction condition (bizarre imagery versus verbal mediation) and type of list (I versus M) was not obtained. Therefore, the results of this study are in agreement with the results of Experiment III in that both experiments failed to reveal any differential performance as a

function of instruction condition (i.e., main effect or interactions).

An analysis was made of the number of words correctly paired, the number of intrusions, and the intrusion rates to determine whether any of these measures differentiated the two instruction conditions. As in the previous experiments, most of the words recalled by Ss using bizarre imagery or verbal mediation were recalled in the correct position. The mean number of misplaced words was .44 and 1.54 for Conditions 1 and 4 respectively; this difference is not significant. The number of intrusions and the error rates did not differ for the two conditions. These findings suggest that the instruction conditions of Experiments III and IV are functionally equivalent in that the performance of Ss instructed to use bizarre imagery did not appear to differ in any way from Ss instructed to use a verbal mediation strategy.

EXPERIMENT V

Paivio (1965) tested the assumption that concrete (C) nouns (e.g., string, tree, coffee) are superior to abstract (A) nouns (e.g., idea, moment, soul) in their capacity to elicit sensory images, and that imagery can mediate the associative connection between items in a pair. Paivio predicted that learning would be better if the concrete noun were in the stimulus position than if it were in the response position because the tendency of the word to arouse mediating imagery should be greater when the word is in the stimulus position. That is, if sensory images can effectively mediate the learning of paired-associate pairs, and if one assumes that the sensory images elicited by the stimulus term are more crucial to paired-associate learning than the sensory images elicited by the re-

sponse term then it follows that C-A lists should be easier to learn than A-C lists.

Paivio predicted that the ease of learning four stimulus-response combinations constructed from concrete and abstract nouns would be in the order C-C, C-A, A-C, A-A. This prediction was supported. Also the ratings for the words on the ease with which they aroused sensory images indicated, as expected, that concrete nouns elicit images more readily than abstract nouns. This study, therefore, strongly suggests that the variable of abstractness is crucial to any consideration of a mnemonic device employing imagery. Moreover, if imagery is responsible for the superior performance on concrete lists, then it should follow that one must predict a greater facilitation of performance for concrete (vs. abstract) words under instructions to employ imagery than under nonimagery instructions. That is, the difference in performance between concrete and abstract lists should be greater under a mnemonic system condition than under a verbal mediation condition.

Method

The design of this experiment was essentially identical to Experiments III and IV in that a bizarre imagery condition and a verbal mediation condition were utilized. The principal difference was in the lists. Four lists of concrete and abstract nouns were constructed by using the 240 (120 concrete and 120 abstract) words listed by Gorman (1961). The four lists of 30 pairs each represented the four possible combinations of peg and response word abstractness (i.e., C-C, C-A, A-C, A-A). The 120 abstract and 120 concrete words were randomly assigned to their four possible positions (2 peg-word lists and 2 response-word lists). The order of the words within any of the resulting paired-associate lists was random with the only restriction that any peg and response word pair did not start with the same letter. A total of 48 Ss (24 per condition) was presented each of the four lists using a proce-

dure essentially identical to that used in Experiments I-IV. Two Ss (one from each instruction condition) were run at the same time. The order of presentation of the four lists was completely counterbalanced.

Results

The mean number of words correctly recalled, disregarding order of recall, for the bizarre imagery condition, was 19.92, 18.25, 16.46, and 8.62 for the C-C, C-A, A-C, and A-A lists, respectively. Mean recall for the verbal mediation condition was 20.58, 17.33, 17.04, and 10.42 for the C-C, C-A, A-C, and A-A lists respectively. A Latin-square design analysis of variance of these scores is presented in Table 4. As expected, word abstractness (C-C versus A-A) had a strong influence on amount recalled. The failure to find a significant superiority of the C-A over the A-C list (locus of abstractness) is not entirely consistent with Paivio's (1965) findings, but this failure to replicate could be due to procedural differences. In the present experiment the dependent measure was the number correctly recalled on the first (and only) recall trial; in the Paivio study the dependent measure was the number correctly recalled on the first four recall trials. It is possible that one trial does not provide sufficient time for a C-A over A-C superiority to be manifested. Since different words were used for the four lists it might be that the words comprising the C-A and A-C lists were of unequal difficulty. That is, even though the words were assigned to the four lists on a random basis, the concrete or abstract words of one list may have been easier than those of another list. Yet the fact that the results of the present experiment are so similar to Paivio's results suggests that this was not the case. As in the Paivio study, the ease of learning the four

TABLE 4
ANALYSIS OF EXPERIMENT V

Source of variance	df	MS	F
Groups (G)	1	13.54	1.53
Sequence (S)	23	61.58	6.99***
Ordinal position (OP)	3	91.83	10.42***
Lists (L)	(3)	(1020.64)	(115.85)***
Abstractness (A)	1	2762.76	318.29***
Locus of A	1	26.04	3.00
A × Locus of A	1	273.13	31.47***
G × S	23	110.59	12.55***
G × OP	3	7.23	—
G × L	(3)	(14.82)	(1.68)
G × A	1	7.50	—
G × Locus of A	1	13.50	1.53
G × A × Locus of A	1	23.37	2.65
Pooled residual	132	8.81	
Total	191		

*** $p < .001$.

lists in the present experiment was in the order C-C, C-A, A-C, A-A. Also, the superiority of the C-A over the A-C list approached significance ($.10 > p > .05$). In any case, the primary intent of the present study was not to replicate Paivio's findings, but to test the notion that instructional conditions could produce differential performance on the lists.

Instruction condition (Groups) and the interactions of instruction condition with the list variables (i.e., Groups × Abstractness, Groups × Locus of Abstractness, and Groups × Abstractness × Locus of Abstractness) failed to reach significance. Thus, these results are consistent with the earlier experiments of this series in failing to demonstrate a differential effect of instructions to use verbal mediation versus instructions to use bizarre imagery. However, there is slight support for Paivio's notion that the tendency of a word to arouse mediating imagery should be greater when the word is in the stimulus position. A comparison of the mean recall for the bizarre imagery and verbal mediation conditions reveals that the bizarre imagery condition was superior (though not significantly so) to the verbal mediation condition only for the C-A list. Also, the dif-

ference in performance between the C-A and the A-C lists was greater for Ss in Group 1 (bizarre imagery) than Group 4 (verbal mediation). Thus, instructions to use imagery slightly increased a difference believed to be caused by the use of imagery.

The significant interaction between Abstractness × Locus of Abstractness (under the heading Lists in Table 4) is a rather interesting finding in that it suggests that increasing the concreteness of the stimulus or the response of an A-A list (i.e., comparing an A-A list with an A-C or C-A list) produces a greater change in performance (facilitation) than the change obtained by decreasing the concreteness of a C-C list (i.e., comparing a C-C list with an A-C or C-A list).

The finding of a significant ordinal position effect is consistent with the earlier experiments in this series. Once again, there was a significant increase in performance as a function of practice. An analysis was made of the number of words correctly paired, the number of intrusions, and the intrusion rates to determine whether any of these measures differentiated the two instruction conditions. None did.

DISCUSSION

Brief Summary of Results

The main findings of the five experiments were as follows: Groups given peg words and the group instructed to link successive items of the list with a bizarre image performed better than a control group given standard free-recall instructions. The groups which differed only with respect to rehearsal instructions had essentially equivalent recall; the groups which differed only with re-

spect to bizarreness (i.e., common versus bizarre images) instructions had essentially equivalent recall. Recall scores on relatively high frequency concrete noun lists (Experiment I), a negative transfer paradigm (Experiment III), a high interference list (Experiment IV), a "mediation" list (Experiment IV), and lists of varying degrees of word abstractness (Experiment V) were not significantly influenced as a function of whether the instructions were to link peg and response words by verbal mediation or by bizarre images. A significant interaction was obtained between Instruction Condition $\times$ Presentation Rate (Experiment II) in that the difference in performance between a group given peg words and instructed to use bizarre imagery and a group given standard free-recall instructions was greater at a 5-second presentation rate than a 2-second presentation rate. There was only very limited support for Paivio's notion that imagery is the factor accounting for the superiority of paired-associate lists consisting of concrete-abstract word pairs over abstract-concrete word pairs since this superiority was only slightly increased by instructing Ss to use imagery to link the words. Finally, Ss given peg words generally showed an improvement with practice while those not given peg words did not. Additional comments are necessary for some of these findings.

Peg Words

First, a procedural question needs to be briefly considered. Rather than have Ss memorize the peg list(s) prior to the presentation of the response lists, Ss were presented with the peg words during both the learning and recall trial. Since this is not the way a mnemonic system is typically used by "mnemonic experts" or

by other investigators of mnemonics (e.g., Smith & Noble, 1965), generalizations from this study may be limited. Yet, despite the interpretational limitations, there are several advantages to the procedure utilized. If Ss in the experimental conditions had been given special training or practice on a system, this experience may have produced differential motivational effects and/or nonspecific transfer effects on the test lists. But, most important, by making the peg list(s) available, the possibility of facilitating recall with a mnemonic system, if, in fact, the effective utilization of a mnemonic system does result in superior recall, is maximized. The argument is that a test of a particular strategy or system should *begin* by utilizing those conditions which are allegedly ideal for the functioning of the system.

Why does the presence of peg words, under the present conditions, facilitate recall? It is apparent, since peg words and response words were consistently paired correctly, that the peg words served as stimuli for the words to be recalled. Since the providing of a stimulus for each response word, in effect, makes the free-recall task a paired-associate task, the problem is explaining why, with the present materials, a paired-associate task should be easier than a free-recall task. Was the presence of the stimulus words solely responsible for the facilitated recall, or were the strategies utilized by the Ss also an important factor? It is possible that the strategy S used to link peg and response words was primarily responsible for the facilitated recall. That is, Ss given peg words and standard paired-associate instructions may only have performed at a level equal to the groups not given peg words. Since a group given standard paired-associate

instructions was not utilized, there is no evidence as to whether the use of imagery or mediation would facilitate performance relative to a control given standard paired-associate instructions.

Other investigators have attempted to determine how mediation instructions affects the learning of paired-associate lists. Although Garskof, Sandak, and Malinowski (1965) were able to facilitate the performance of Ss given two successive lists (A-B, A-D paradigm) by instructing Ss to mediate, it was not clear from their study, since they only reported the mean performance on the two lists combined, whether the instructions facilitated performance on both the A-B and A-D list or merely on the A-D list. The present concern is not with how instructions to mediate influence transfer; the present concern is whether instructions to mediate facilitate the acquisition of a single list. Martin and Dean (1966) were unable to facilitate first-list learning by giving mediation instructions; McNulty (1966) was unable to facilitate the performance of motivated Ss on a single list by instructing them to mediate. In light of the absence of a clear demonstration that instructions to mediate facilitate the acquisition of a single list for normal, motivated Ss, and the failure of the different instruction conditions of the present experiments to influence performance, it is likely that instructing Ss to use a particular strategy has, at best, only a slight effect on the acquisition of a single paired-associate list.

Before generalizing that providing Ss with a set of stimuli (peg list) facilitates recall it would seem necessary to have Ss learn other kinds of lists. For example, would a list which required a considerable amount of response learning (such as a low-

meaningfulness nonsense-syllable list) be facilitated by the presentation of a peg-word list? The point is that in the present study response learning was minimal because the words were of high meaningfulness, therefore, the associative stage was the principal constituent of the learning. However, it might be expected that providing Ss with a peg list would not facilitate recall if the material to be learned required a considerable amount of response learning. An interaction between meaningfulness and the presence or absence of a peg list should occur such that the superiority of the peg-list condition relative to the control would be greater for lists of high meaningfulness than for low meaningfulness. In fact, evidence from the present study indicates that this interaction might be obtained even when words are used for the peg and response lists. Although the appropriate control group was not run, the low mean recall on the abstract-abstract list (8.62 and 10.42 for Conditions 1 and 4 respectively) of Experiment V strongly suggests, to the extent that abstractness and meaningfulness are correlated, that the superiority of a peg-list condition relative to a control not given a peg list decreases as the meaningfulness of the peg and response words decreases.

If, as in the present study, providing Ss with a stimulus facilitated recall of high meaningful word lists, it should follow that other attempts to "provide" or "define" the stimulus might also result in facilitated performance. Thus, instructing Ss to use contextual objects or other items in the list as stimuli should result in facilitated recall. There is also reason to suspect that the nature of the stimuli might determine whether or not supplying the stimuli for the recall of the response words facilitates

performance, that is, compare the performance on the concrete-abstract and abstract-abstract lists of Experiment V. Also, it has yet to be established whether numbers or the alphabet can serve as effective stimuli for recall when the materials to be recalled are not specifically designed to conform to a numerical or alphabetical sequence. If numbers or the alphabet can serve as effective stimuli, a peg list is, of course, readily available for the facilitation of recall.

Besides the present study, there is apparently only one other reported instance of facilitated "free-recall" learning resulting from instructing Ss on a particular strategy. Tulving (1962) demonstrated that instructing Ss to use an alphabetic organization of the items produced significantly superior recall relative to a control group given standard free-recall instructions. The Ss in the instructed condition were told to make associations between each letter of the alphabet and the response word having that initial letter. Each response word started with a different letter. Thus, Ss in effect were given a peg list (since the alphabet is highly available) and told to make associations between each peg and response word. The present study extends the Tulving finding in that the peg lists and response lists for the present study were randomly selected from a pool of words and not limited to the alphabet or to sets of words all having different initial letters.

Linking Successive Items

There are at least two ways to interpret the finding that instructing Ss to link successive items of the list with a bizarre image resulted in superior recall relative to a control. Deese (1960) maintains that the inter-item associative strength of the

words in the list determines amount of recall because words that are associated elicit each other at recall. Underwood and Schulz (1960) suggest that response recall is a matter of increasing the strength of the response words to the contextual situation so that the response strengths of items in the list are greater than other responses that are not in the list. The major difference between the two positions appears to be whether the emphasis is on the stimulus eliciting the response or increasing the response strength so that the response can be elicited by contextual stimuli. That is, Underwood and Schulz emphasize the *frequency* of elicitation of each item in the list. Deese emphasizes the *associational* connections among items in the list. Since the instructions to associate successive items resulted in superior recall it might appear that Deese's position was unequivocally supported. However, it can also be maintained that instructions to associate items increase the frequency of elicitation of the items in that Ss are "forced" to rehearse earlier items in order to make associations among the items. Fortunately, it appears a differential prediction can be made from the above two positions regarding the order of recall of the items. It follows from Deese's point of view that items which have been associated during learning should occur together in recall if, in fact, words that are associated *elicit* each other during recall. However, if frequency is the crucial variable then there is no reason to suspect that items occurring together during learning should necessarily occur together at recall since no prediction regarding order of recall follows from the frequency principle. Thus, if an increase in inter-item associative strength was responsible for

the facilitated recall of the group instructed to link successive items during learning, the order of recall for this group should, relative to a control, closely approximate the order of presentation of the list.

To test the above prediction the order of recall of the group told to link successive items and the group given standard free-recall instructions (Experiment I) was examined. The number of pairs of words occurring successively in recall that were presented successively during the learning trial was counted for these groups. The order of the words within the pair was ignored, that is, if the first and second words presented were recalled in reverse order (second word first and first word second) this was still counted as a cluster. An *S* having perfect clustering would have one less cluster than the number of words correctly recalled. For each *S* in the two conditions the total number of clusters was obtained for each list and divided by the total number of words correctly recalled to yield a cluster rate. A mean cluster rate was obtained for each *S* over all three lists. The mean for the *Ss* in the group instructed to link successive items was .506; the mean for the *Ss* in the control was .377. This difference was significant ($F = 5.09$, $df = 1/34$, $p < .05$). Thus, this analysis supported Deese's notion that inter-item associative strength determines amount recalled in that the order of recall was shown to reflect the associative connections that *Ss* were instructed to make during the learning trial. It is of course possible that the Underwood and Schulz frequency notion could be extended to include a prediction regarding order of recall. It seems unlikely, however, that this could be done without ascribing an eliciting function to the words in the list.

Rehearsal

Although the instructions to rehearse produced a slight, though insignificant, facilitatory effect, the interpretation of these results must be extremely guarded because there is some doubt as to whether or not the instruction conditions actually produced a difference in rehearsal. If the group told to rehearse previously formed images whenever possible did not have time to do so, then the rehearsal variable was not manipulated in the present study. Unfortunately, even though the *Ss*, when queried about whether they were able to follow their special instructions, reported that they were able to follow the instructions, no measure was obtained of the extent to which they rehearsed previous items. In short, no definitive statement regarding the influence of rehearsal can be made.

Presentation Rate

Investigators have reasoned that if *Ss* use mediation, optimal performance should be obtained with an anticipation interval which is sufficiently long to give the *Ss* an opportunity to utilize mediators. Some investigators (e.g., Runquist & Marshall, 1963; Schulz & Lovelace, 1964) have obtained results which support this notion and others (e.g., Schwenn & Underwood, 1965; Spear, Mikulka, & Podd, 1965) have not. In light of the conflicting findings and the lack of an apparent reason for the conflicting findings, no general conclusion is warranted regarding the presence or absence of an optimal presentation rate when *Ss* use mediation. It should be noted that the above studies utilized transfer paradigms (A-B, A-B'; simulated A-B, A-B'; or A-B, C-D, A-D where a preestablished association between B and C was inferred

from association norms). Another approach to the study of presentation rate and mediation would be to have Ss learn a single list with different presentation rates under mediation or nonmediation instructions. It seems plausible that the failure to demonstrate a reliable facilitory effect on a single list as a function of mediation instructions may be due to the failure to use slow presentation rates (5 seconds or more). In any case, for the conditions of the present study (Experiment II), the results clearly indicate that for Ss using peg lists a presentation rate of 5 seconds per item is more "nearly optimal" than a presentation rate of 2 seconds per item.

Imagery as an Association Aid

The present series of studies afford little support for the view that imagery serves as an effective associational aid in that instructing Ss to use imagery did not facilitate performance relative to the group (verbal mediation) not instructed to use imagery. In Experiment V, even though the difference between the concrete-abstract and abstract-concrete list was greater for Ss in the imagery instruction condition, the failure to obtain a significant interaction between the locus of abstractness (stimulus or response) and instruction condition suggests that either the utilization of imagery does not facilitate performance under these conditions, or that the utilization of imagery cannot be effectively manipulated by instructions. Thus, Experiment V does not provide convincing support for the notion that the superiority of concrete-abstract lists to abstract-concrete lists is due to imagery.

Imagery has also been offered as an explanation for the superiority of noun-adjective lists to adjective-noun lists, as nouns are believed to elicit

more images than adjectives. The rationale for the superiority of noun-adjective lists over adjective-noun lists is the same as for the superiority of concrete-abstract lists over abstract-concrete lists (Lambert & Paivio, 1956). A recent study by Perrino (1965), however, gives considerable doubt to the notion that imagery is involved in producing this phenomenon. Perrino, among other things, embedded disyllable nonsense words in grammatical sentence frames by using English function words and verbs. The nonsense words were embedded in adjective and noun positions. After Ss were given trials in which they learned to recognize the grammatical strings, they were given a paired-associate list in which the nonsense pairs, now presumed to have acquired form class (noun or adjective) membership, served as the stimuli and responses. The pairs having noun-adjective form class were learned more rapidly than the adjective-noun pairs for the first six trials. Although the absolute differences were not large, grammatical structure appears to have been responsible for producing the effect. Since there seems to be little justification for maintaining that the nonsense disyllables differed on imagery, there is little reason to suspect that imagery functioned as an associational aid to produce these differences. Thus, if imagery functions as an associative aid, it is not the sole factor producing the noun-adjective, adjective-noun effect and presumably the concrete-abstract, abstract-concrete effect.

Nonspecific Transfer

The finding of positive nonspecific transfer for Ss receiving peg words and the failure to find positive nonspecific transfer (warm-up and/or learning to learn) for Ss not re-

ceiving peg words is consistent with the findings of other investigators. Since Ss receiving peg words in actuality are being given a paired-associate task while those Ss not given peg words have a free-recall task, the issue is why nonspecific positive transfer is obtained using a paired-associate procedure and not with a free-recall procedure. With respect to paired-associate learning, the finding of positive transfer is consistent with other studies (e.g., Postman & Schwartz, 1964; Thune, 1951). Yet for free recall the situation is more complex. Dallett (1963) points out that learning to learn is difficult to obtain when only a few learning trials per list are given. Since only one trial was given per list in the present study, the failure to obtain learning to learn or warm-up effects for those groups not given peg words is consonant with the findings of most other investigators (e.g., Deese, 1957; Murdock, 1960). Yet, Murdock (1962) gave each S 20 lists per session, one trial per list, and four different sessions, and found improvement across sessions. However, the effect was small for all six of his conditions and significant for only four of the six conditions. If only the first recall trial is considered, Tulving, McNulty, and Ozier (1965) and Dallett (1963) were also unable to demonstrate a learning to learn effect. Learning to learn has been obtained in free recall when Ss have been given repeated trials on the same list (Dallett, 1963; Meyer & Miles, 1953; Tulving et al., 1965).

Implications of This Research

The present study strongly suggests that the utilization of a well-memorized peg list can produce a marked facilitation, relative to a control, in the number of words that Ss are able to remember. Yet this does not neces-

sarily mean that memory systems should be advocated. The decision to adopt a mnemonic system approach to the problem of how best to memorize verbal material should depend, hopefully, on more considerations than the demonstration that a particular approach can, under some circumstances, result in facilitated recall. Although the present study does not provide an unequivocal answer to the question of whether to adopt a mnemonic system approach, it gives at least partial answers for many of the considerations.

There are compelling reasons for seriously considering a mnemonic system approach as a means for facilitating memory. The present study suggests that: (a) The amount of recall of concrete and perhaps abstract words is increased when a peg list is used; (b) the correct serial ordering of recall seems to be an inherent result of a peg list approach; (c) material in the middle of a list does not appear to be any more difficult to learn than material at the ends of the list. The question of whether the recall of abstract words is facilitated when a peg list is utilized requires further elaboration. Although a group given standard free-recall instructions (no peg list) was not employed in Experiment V, an estimate of the number of words such a group would have recalled can be made by utilizing Murdock's (1960, p. 231) formula. A recall of approximately 15 words would be expected from a list of 30 words presented at a rate of one word every 5 seconds. Yet the two groups given peg lists consisting of concrete words had mean recalls of 18.25 and 17.33 for the list of 30 abstract words. This evidence, albeit tenuous, suggests that the presence of a peg list can also facilitate the recall of abstract words.

Several facts argue against the use

of a peg-list approach to memorizing verbal material. Obviously, a certain amount of time and effort is necessary in order to learn the peg list. Thus, unless the peg list is used to learn many lists, it would seem doubtful whether the approach would result in an overall reduction in the amount of time required to memorize the material. Yet performance is likely to decrease as a function of the number of times the peg list was previously used. That is, although there is improvement with practice if a different peg list is used for each response list, the time required to master a peg list generally precludes memorizing more than one list. Given that only one peg list is used, the results of Experiment III suggest that performance will decrease with each succeeding list. Thus, in the case in which lists of unrelated words are to be recalled, the effectiveness of a peg-list approach would seem to depend on whether Ss using a peg list continue to be superior to control Ss when numerous (say several hundred) lists are learned.

A final note of caution seems to be in order. Generalizations from the present study to practical learning situations should be guarded for at least two additional reasons. First, there is little evidence concerning the effectiveness of a mnemonic system approach when "organized" or complex material is to be memorized. Second, the present study does not afford any information regarding the long-term retention of material learned by means of a peg-list approach. Yet, obviously, when many lists are learned by means of the same peg list the retention of these lists will be markedly hindered by retroactive and/or proactive interference. The extent of this interference relative to a group not utilizing peg lists remains undetermined.

SUMMARY

A series of five experiments was conducted to determine under what circumstances, if any, mnemonic systems facilitated recall, the elements of the system responsible for the "facilitated" recall, and the relationships between mnemonic systems and some other variables known to affect learning. The basic component of a mnemonic system is generally a list of peg words. During learning one makes an association between each word to be memorized and a peg word. Then for recall, the peg word, which is readily available, "elicits" the word to be recalled. In Experiment I the recall of a list of 40 words was compared for four groups provided with a peg list and three groups not provided with a peg list. The four groups receiving peg lists differed with respect to the kind of association they were asked to make to connect each peg and response word. Two groups were told to use bizarre imagery. One group was told to use common imagery. The fourth group was told to use a form of verbal mediation. One of the groups told to use bizarre imagery was instructed to rehearse previous images; the other group was told not to rehearse. Of the three groups not given peg words, one group was told to make an image of each response word. The second group was instructed to link successive response words with a bizarre image. The third group, a control, received standard free-recall instructions. In Experiment II a 2-second and a 5-second presentation rate were used to determine if there was an optimal presentation rate for Ss utilizing mnemonics. In Experiments III-V a bizarre imagery group and a verbal mediation group were utilized to assess whether these groups would perform differentially on a negative

transfer paradigm (Experiment III), a high interference or a "mediation" list (Experiment IV), and word tests of varying degrees of abstractness (Experiment V).

The main findings of the five experiments were: Groups given peg words and the group instructed to link successive items of the list with a bizarre image performed better than the control given standard free-recall instructions. Rehearsal and "bizarreness" instructions had no apparent effect on performance. Recall scores on relatively high frequency concrete noun lists (Experiment I), a negative transfer paradigm (Experiment III), a high interference list (Experiment IV), a "mediation" list (Experiment IV), and lists of varying degrees of word abstractness (Experiment V) were not significantly influenced as a function of whether the instructions were to link peg and response words by verbal mediation or by bizarre images. A significant interaction was obtained between instructions and presentation rate (Experiment II) in that the difference in performance between a group given peg words and instructed to use bizarre imagery and a group given standard free-recall instructions was greater at a 5-second presentation rate than a 2-second presentation rate. The Ss given peg words generally showed an improvement with practice while those not given peg words did not.

The nature of the stimulus in recall, the relationship between the presence of peg words and recall for materials of varying degrees of meaningfulness, a frequency versus an inter-item associative strength interpretation of free-recall learning, imagery versus verbal mediation, rehearsal, presentation rate, imagery as an associational aid, and nonspecific

transfer were the problems briefly discussed.

REFERENCES

- BALABAN, A. Über den Unterschied des logischen und des mechanischen Gedächtnisses. *Zeitschrift für Psychologie*, 1910, **56**, 356-377.
- BATTIG, W. F., BROWN, S. C., & NELSON, D. Constant vs. varied serial order in paired-associate learning. *Psychological Reports*, 1963, **12**, 695-721.
- BOUSFIELD, W. A., WHITMARSH, G. A., & ESTERSON, J. Serial position effects and the "Marbe effect" in the free recall of meaningful words. *Journal of General Psychology*, 1958, **59**, 255-262.
- BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology*, 1962, **63**, 409-412.
- DALLETT, K. M. Practice effects in free and ordered recall. *Journal of Experimental Psychology*, 1963, **66**, 65-71.
- DEESE, J. Serial organization in the recall of disconnected items. *Psychological Reports*, 1957, **3**, 577-582.
- DEESE, J. Frequency of usage and the number of words in free recall: The role of associations. *Psychological Reports*, 1960, **7**, 337-344.
- DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized material. *Journal of Experimental Psychology*, 1957, **54**, 180-187.
- FURST, B. *The practical way to a better memory*. New York: Fawcett World Library, 1957.
- GARSKOF, B. E., SANDAK, J. M., & MALINOWSKI, E. W. A. Controlling the "fate" of first list associates. *Psychonomic Science*, 1965, **2**, 315-316.
- GORMAN, A. M. Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, 1961, **61**, 23-29.
- GRANT, D. A. The Latin square principle in the design and analysis of psychological experiments. *Psychological Bulletin*, 1948, **45**, 427-442.
- LAMBERT, W. E., & PAIVIO, A. The influence of noun-adjective order on learning. *Canadian Journal of Psychology*, 1956, **10**, 9-12.
- MARTIN, R. B., & DEAN, S. J. Reported mediation in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 23-27.

- McNULTY, J. A. The effects of "instructions to mediate" upon paired-associate learning. *Psychonomic Science*, 1966, **4**, 61-62.
- MEYER, D. R., & MILES, R. C. Intralist-interlist relations in verbal learning. *Journal of Experimental Psychology*, 1953, **45**, 109-115.
- MURDOCK, B. B. The immediate retention of unrelated words. *Journal of Experimental Psychology*, 1960, **60**, 222-234.
- MURDOCK, B. B. The serial position effect of free recall. *Journal of Experimental Psychology*, 1962, **64**, 482-488.
- NUTT, R. H. *How to develop a good memory for names, faces, and facts*. New York: Simon & Schuster, 1941.
- PAIVIO, A. Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 32-38.
- PERRINO, C. An evaluation of the effect of grammatical structure on learning. Unpublished master's thesis, Northwestern University, 1965.
- POSTMAN, L., & SCHWARTZ, M. Studies of learning to learn: I. Transfer as a function of method of practice and class of verbal materials. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 37-49.
- ROTH, D. M. *Roth memory course*. Santa Monica: Motivation, 1961.
- RUNQUIST, W. N., & FARLEY, F. H. The use of mediators in the learning of verbal paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 280-285.
- RUNQUIST, W. N., & MARSHALL, M. A. Transfer, synonymity, and anticipatory interval in paired-associate verbal learning. *American Journal of Psychology*, 1963, **76**, 281-286.
- RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, Contract N8 ONR-66216, Office of Naval Research.
- SCHULZ, R. W., & LOVELACE, E. A. Mediation in verbal paired-associate learning: The role of temporal factors. *Psychonomic Science*, 1964, **1**, 95-96.
- SCHWENN, E., & UNDERWOOD, B. J. Simulated similarity and mediation time in transfer. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 476-483.
- SMITH, R., & NOBLE, C. E. Effects of a mnemonic technique applied to verbal learning and memory. *Perceptual and Motor Skills*, 1965, **21**, 123-134.
- SPEAR, N. E., MIKULKA, P. J., & PODD, M. Transfer as a function of time to mediate. *Journal of Experimental Psychology*, 1966, **72**, 40-46.
- THUNE, L. E. Warm-up effect as a function of level of practice in verbal learning. *Journal of Experimental Psychology*, 1951, **42**, 250-256.
- TULVING, E. The effect of alphabetical subjective organization on memorizing unrelated words. *Canadian Journal of Psychology*, 1962, **16**, 185-191.
- TULVING, E., McNULTY, J. A., & OZIER, M. Vividness of words and learning to learn in free-recall learning. *Canadian Journal of Psychology*, 1965, **19**, 242-252.
- TWEDT, H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *Journal of Experimental Psychology*, 1959, **58**, 111-116.
- UNDERWOOD, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first list learning. *Journal of Experimental Psychology*, 1951, **42**, 44-54.
- UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychological Bulletin*, 1956, **53**, 84-95.
- UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.
- WALLACE, W. H., TURNER, S. H., & PERKINS, C. C. *Preliminary studies of human information storage*. Signal Corps Project No. 1320, 1957, Institute for Cooperative Research, University of Pennsylvania.

APPENDIX A

TABLE A1

PEG WORDS AND RESPONSE WORDS FOR EXPERIMENTS I-III

List I	List II	List III
Peg—Response	Peg—Response	Peg—Response
wheel-kitten	beet-limousine	hailstone-barn
flea-actor	puddle-village	bread-revolver
armor-closet	coffee-pillow	camel-tack
pearl-gas	bungalow-oyster	jellyfish-donkey
onion-balloon	pine-straw	city-diamond
pineapple-blood	atom-custard	gasoline-flannel
dime-brick	bracelet-ether	cranberry-turpentine
moss-peach	cradle-garbage	aluminum-flask
bed-whale	canary-buckle	cheese-fur
alley-tomato	spinach-doughnut	earthworm-vinegar
dirt-honey	lint-bandage	dagger-jewel
germ-spear	chestnut-needle	barrel-feather
baseball-cork	mouse-tobacco	asparagus-head
boulder-sewer	sardine-belly	dome-stadium
teeth-forest	baton-milk	fang-sunflower
dandelion-knuckle	moccasin-auditorium	grape-lard
enamel-nutshell	ammonia-bicycle	hog-gorilla
armpit-manure	olive-lizard	beak-grasshopper
tongue-derby	skin-ape	carrot-daffodil
minnow-asphalt	pickle-tweezer	lawn-money
button-cabbage	harpoon-rhinestone	zoo-corn
ski-fire	knob-ivy	ice-anchor
lemon-spool	walrus-seaweed	dish-moon
fence-diary	booze-globe	bacon-scissors
crown-telephone	elephant-brass	paste-ocean
salt-chalk	fishhook-crumb	measles-cigarette
icicle-crystal	ivory-saucer	skull-linen
cave-eel	frost-padlock	glue-pill
snow-mansion	milk-butter	egg-night
rocket-goat	adobe-grass	pot-bone
jug-napkin	cucumber-helmet	sugar-platter
sulphur-pup	heart-drum	sauerkraut-badger
chocolate-pollen	velvet-horse	tunnel-cabin
notebook-pin	stone-monkey	rice-collar
cinnamon-rod	diaper-coal	pear-rattlesnake
hallway-cherry	gnat-skunk	radio-bean
hairpin-lantern	pencil-pony	lips-capsule
ginger-freckle	thimble-cigar	mustard-sheep
hatchet-grapefruit	hospital-cauliflower	tar-snail
rabbit-apple	eyeball-knife	silk-banana

TABLE A2
WORD LISTS FOR EXPERIMENT IV

List I		List M	
Peg	Response	Peg	Response
cheese	black	platter	saucer
window	hot	silk	chamois
mountain	apple	ether	cigar
short	foot	cheese	mustard
eagle	sour	diamond	aluminum
spider	blue	cave	alley
lion	smoke	anchor	boulder
man	tiger	paste	tar
smooth	drink	diaper	cigarette
cabbage	home	bean	ivy
white	tall	ginger	clove
stem	head	asphalt	coal
whistle	lamb	lard	butter
tobacco	chair	tomato	flannel
bread	army	rod	hallway
king	butter	nail	ocean
fruit	queen	cucumber	straw
table	crackers	chestnut	freckle
cold	thread	button	doughnut
yellow	water	dagger	fishhook
river	train	hog	garbage
sheep	moth	eel	lizard
mutton	food	icicle	frost
hammer	light	auditorium	mansion
whiskey	nurse	buckle	napkin
cottage	web	cork	balloon
priest	bird	city	telephone
sweet	rug	sheep	peach
butterfly	hill	honey	sugar
needle	rough	tweezer	armor
soldier	door	grapefruit	sauerkraut
house	woman	velvet	kitten
hand	fast	baton	rattlesnake
dark	round	atom	crumb
earth	wool	puddle	gasoline
carpet	flower	rabbit	ape
eating	church	hailstone	bone
doctor	stars	crystal	turpentine
slow	nail	beak	needle
moon	lake	barrel	cradle

APPENDIX B

INSTRUCTIONS FOR EXPERIMENT I

General Instructions for Experiment I

This is an experiment on memory for words. Three lists of 40 words each will be read to you on this tape recorder. The words will be read at a rate of one word every 5 seconds. Each word will be read once; each list will be read once. Following the presentation of a list you will be asked to write as many of the words as you can remember on an answer sheet. As nearly as possible write down the words in the same order that they were read to you. That is, write the first word of each list in Blank 1 of your answer sheet, the second word in Blank 2, etc. If you do not remember the position of the word in the list make a guess as to its position. Your ability to recall the words is of primary interest, remembering the position of the word in the list is secondary.

I am going to pass out some five-page handouts. Do not open your handout until you are told to do so. Pages 3-5 of your handout are answer sheets. Page 3 is to be used for the first list, Page 4 for the second list, and Page 5 for the third list. On the second page of your handout, you will find special instructions explaining how you are to learn the lists of words that will be read to you. The purpose of this experiment is to determine the effectiveness of different strategies for the memorizing of words, so it is extremely important that you utilize, to the best of your ability, the strategy that is explained to you on Page 2 of your handout.

If you have any questions, now is the time to ask them. However, it will not be possible to answer questions concerning the particular strategies to be employed because some of you are being asked to employ different approaches. Therefore, questions concerning one strategy would only confuse the users of a different strategy. You will have to study Page 2 until you understand what is expected of you. Since some of the strategies are more complicated than others, we will wait until everyone assures me they understand

their approach before starting the tape recorder.

Special Instructions for Group 1 Experiment I

Forty words have been written on your answer sheet to aid you in the memorizing of the words to be read on the tape recorder (tape words). Basically, your task is to make a bizarre mental picture using each answer-sheet word and each taped word. That is, you are to try to form a bizarre visual image linking the first taped word (i.e., the first word read) and the first answer-sheet word (i.e., the first word on your answer sheet); then you are to try to form a bizarre visual image linking the second taped word and the second answer-sheet word; then you are to try to form a bizarre visual image linking the third taped word and the third answer-sheet word, etc. Thus, after a list of taped words has been read, you should have formed (ideally) visual images linking each of the 40 taped words with one of the 40 answer-sheet words.

It is essential that you form a *bizarre* mental picture to link each answer-sheet and taped word because *bizarre* images tend to be easier to remember. For example, if *automobile* was the first word printed on your answer sheet and *saltshaker* was the first taped word, you might imagine a huge *saltshaker* driving an *automobile*. This bizarre mental picture should be rather easy to remember. Also, you should make no attempt to rehearse previously formed images.

To recall the list of 40 taped words, you are to use the answer-sheet words. Each answer-sheet word (e.g., *automobile*) should evoke the bizarre image linking the taped word and answer-sheet word (e.g., a huge saltshaker driving an automobile), thus providing you with the taped word (*saltshaker*).

Special Instructions for Group 4 Experiment I

Forty words have been written on your answer sheet to aid you in the memorizing

of the words to be read on the tape recorder (taped words). Basically, your task is to link each answer-sheet word and each taped word by a verbal connection. That is, you are to try to verbally link the first answer-sheet word (i.e., the first word on your answer sheet) and the first taped word (i.e., the first word read); then you are to try to verbally link the second answer-sheet word and the second taped word; then you are to try to verbally link the third answer-sheet word and the third taped word, etc. Thus, after the taped words have been read, you should have (ideally) each of the 40 taped words verbally linked with one of the 40 answer-sheet words.

You are free to link each answer-sheet word to a taped word with any verbal connection that seems appropriate. For example, if you want to link the answer-sheet word *dog* with the taped word *nine*, you may use the word *cat* to connect the two words. Then during recall the answer sheet word *dog* should elicit *cat* and *cat* should elicit *nine*. If you want to link the answer-sheet word *automobile* with the taped word *saltshaker*, you might construct a sentence using *saltshaker* and *automobile* as key words (e.g., The saltshaker is on the automobile.). Also, you may find it unnecessary to provide a verbal link for some of the answer-sheet and taped word pairs, that is, the two words may just "seem to go together."

To recall the list of 40 taped words, you are to use the answer-sheet words. Each answer-sheet word should elicit a taped word because you have connected each answer-sheet word with a taped word. For example, *automobile* should remind you of the sentence, "The saltshaker is on the automobile." and thus provide you with the peg word *saltshaker*.

Special Instructions for Group 6 Experiment I

Basically your task is to make a bizarre mental picture to link successive words in the list. For example, if *automobile*, *saltshaker*, and *dog* were the first, second, and third words respectively of the list, your task would be to form a bizarre image to link these three words. A possible bizarre image might be a huge *saltshaker* driving an *automobile* shaped like a *dog*. In a similar manner, you are to try to form bizarre visual images to link the remaining words of the list. The number of words you include in any one image is optional. At one extreme, you might form 20 images, each containing two words; at the other extreme, you might form one bizarre image containing all the words. It is likely that forming 8-12 images (3-4 words per image) will result in the highest recall. Also, it is not necessary to have any word represented in two images. That is, if your first image links words 1, 2, 3, and 4, your second image should link words other than words 1, 2, 3, and 4 (e.g., words 5, 6, and 7). Thus, after the words have been read, you should have included (ideally) each word in one of your bizarre mental pictures.

To recall the list of 40 words, you need recall only the bizarre images that you have formed. It is expected that recalling bizarre images will be a relatively easy task. That is, one is not likely to forget a mental picture of a huge *saltshaker* driving an *automobile* shaped like a *dog*.

Special Instructions for Group 7 Experiment I

You are free to learn the lists of words in whatever way you feel will result in the best performance. That is, adopt the strategy which is best for you.

(Received November 4, 1966)

